NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

MONITORING AGENCY DOCUMENT NO: AFCRC
ASTIA DOCUMENT NO: AD
CONTRACT AF 61 (052)-342

SUMMARY REPORT

SPEECH ANALYSIS AND SYNTHESIS

C. G. M. Fant
Royal Institute of Technology
Stockholm / Sweden

January 31, 1961

62-1-5

"The research reported in this document has been sponsored by, or in part by Air Force Research Division of the AIR RESEARCH AND DEVELOPMENT COMMAND, UNITED STATES AIR FORCE, through its European Office."

MONITORING AGENCY DOCUMENT NO: AFCRC
ASTIA DOCUMENT NO: AD
CONTRACT AF 61(052)-342

SUMMARY REPORT

SPEECH ANALYSIS AND SYNTHESIS

C.G.M. Fant

Royal Institute of Technology
Stockholm / Sweden

January 31, 1961

"The research reported in this document has been sponsored by, or in part by Air Force Research Division of the AIR RESEARCH AND DEVELOPMENT COMMAND, UNITED STATES AIR FORCE, through its European Office."

TABLE OF CONTENTS

TABLE OF CONTENTS

STL-QPSR-1/1960 (July-September)

TABLE OF CONTENTS

STL-QPSR-2/1960 (October-December)

INTRODUCTION

The present annual summary report No. 1, covering the period 1 December 1959 to 31 December 1960, comprises the following condensed summary and in addition the Speech Transmission Laboratory Quarterly Progress and Status Report Nos. 1 and 2 for 1960 which contain the detailed information and are included as a reference.

SUMMARY AND CONCLUSIONS

I. Speech Analysis

A. Spectrum sampling instrumentation

The design of a 51-channel spectrum analyzer for periodic synchronous sampling and visual display of intensity-frequency sections of connected speech is completed in so far as the filterbank is concerned.

A 6-channel working model of the system intended for closed loop analysis has been constructed and constitutes at present our most accurate device for taking spectral sections. The 51-channel spectrometer which should be completed in about 18 months' time will allow a much greater speed of analysis thus minimizing the routine work in taking the primary data.

An experimental study of the smoothing requirements for analysis of random noise type sounds has been undertaken. Recommendations for improving the function of the sectioning device in the Sona-Graph analyzer have been made.

B. Formant frequency measurements

Systematic errors and average spread in taking formant frequency data from the Sona-Graph have been investigated. Time-frequency-intensity spectrograms provide almost the same accuracy as sections and the systematic error defined as the deviation of measurements from the true pole frequency of vowels is generally less than 50 c/s. This is also the magnitude of the average spread among investigators. The increase of systematic errors and spread as a function of increasing F_0 has been studied.

The work on automatic tracking of formant frequencies has been concentrated to the detection of F_1. Inverse-filtering comprising minimum selection within a bank of anti-resonance filters did not prove to be sensitive enough. All methods tried including moment-weighting, phase detection, and zero-crossing counting have definite limitations at high voice fundamental frequencies. Since these methods do not appear to have any substantial advantages compared with the method of peak-picking from a filterbank it is considered best to concentrate future work on more thorough study of the latter method and specifically how to avoid errors due to a high F_0 and extra formants. The necessary logics for accomplishing this control may be developed according to analysis-by-synthesis concepts.

C. Pole-zero matching techniques

Various analog and numerical methods of synthesizing spectral functions from an inventory of maximally five poles and four zeros have been investigated.

The methods have been applied to fricatives and vowels. It was found that an approximation in terms of two poles and one zero gave a reasonably good fit for fricatives as judged by corroborative synthesis experiments. The vowel studies have been directed towards a pole-zero match of voice sources.

D. Voice source studies

Studies of the vocal source time function and spectrum have been undertaken by means of inverse-filtering techniques. Results support earlier observations on the relative increase in the spectrum level at high frequencies at increasing voice efforts.

The minimum often seen at about 800 c/s in vowel spectra is a typical voice source feature and reflects the duration of the base length of the vocal pulses.

Further systematic studies are needed especially in connected speech.

E. Voice fundamental frequency tracking

The following three methods for increasing the relative level of the voice fundamental before performing the frequency measurements may be recommended as a result of our work:

1. Rectified single side-band, LP 1000 c/s
2. Full-wave rectified band, BP 300-2000 c/s
3. Continuous tracking and rejection of F1 by inverse-filtering

The tendency of a pitch frequency meter to synchronize on overtones may be reduced by means of an independent measurement of F_0 within three separate F_0-channels spaced less than one octave apart. A minimum selector picks the lowest F_0 as the appropriate parameter value. This system supplements an optimal degree of low-pass filtering after the pre-emphasis stage.

F. Automatic identification of sound features

Our studies of schemes for the detection of voicing have led us to recommend a criterion of high energy in the integrated or low-pass filtered speech wave compared with the average for the speaker in this band. The commonly used method of normalizing with regard to the intensity of the high-pass filtered speech of the same sample did not give equally satisfactory results. A useful supplementary criterion in addition to the absolute level of the low-frequency energy is the frequency of zero-crossings in the speech wave which is always less for voiced sounds than unvoiced sounds.

A detection of natural boundaries in connected speech has been attempted on the basis of a running measure of the sum of the absolute values of the changes in the short time spectral intensity.

These changes are determined from a few broad bands covering the spectrum. This running measure shows clear peaks at the boundaries between sound segments as seen from spectrograms.

G. Structural classification of Swedish phonemes

The distinctive feature theory as proposed by Jakobson, Fant, and Halle has been applied to the Swedish phoneme system. The discussion of the nature of the distinctions and their acoustic correlates applies to most languages. A note on the acoustic structure of syllables is included.

II. Speech Synthesis and Speech Perception

A. Confusion among vowels following low-pass and high-pass filtering

Articulation scores for the identification of natural and synthetic vowels as a function of the cutoff frequency of low-pass and high-pass filtering have been determined. Synthetic vowels gave approximately the same results as natural vowels and were equally intelligible.

B. Vowel synthesis

A fairly extensive recording of synthetic vowels of systematically varied formant frequencies has been made.

C. F-pattern approximations of voiced stops and fricatives

A few pilot experiments on the simulation of voiced stops and so-called voiced fricatives on the basis of vowel-like stimuli have been undertaken and the results have been discussed with respect to the Haskins Laboratories data.

D. Continued work on the synthesis of connected speech

Our work on the synthesis of connected speech by means of a parametric control of series type synthesis circuitry has continued. We find the results promising enough to motivate a new project devoted to the development of a complete analysis-synthesis vocoder system along these lines. A high intelligibility and naturalness should be obtainable at an upper limit of 1200 bits per second data transmission rate. At present, however, we do not possess the suffucient funds for the realization of these plans in addition to the other scopes of our present contracts.

III. Speech Production

A. X-ray demonstration film

An X-ray cinefilm illustrating coarticulation effects in human speech has been prepared by H.M.Truby with partial support from the contract.

B. Studies of nasalization

In connection with a project on the study of cleft palate speech in which we have participated there has been the opportunity to study by cineradiographic methods the movements of the soft palate and the associated effects in the spectrographic picture. A substantial part of the study has been devoted to normal speakers. It is found that the movements of the soft palate show a tendency of constant speed independent of the speaking rate. The influence of the nasal coupling on the speech spectrum is noticeable only at coupling areas in the velopharyngeal passage of more than 10 mm^2. The movements of the soft palate from the closed state to the open state begin with a downward movement during which the coupling area is negligible. At very large degrees of nasal coupling the second formant may be attenuated more than the first formant.

LIST OF ILLUSTRATIONS

STL-QPSR-1/1960 page

STL-QPSR-1/1960 page

STL-QPSR-2/1960

SCIENTIFIC PAPER PUBLISHED ON THE CONTRACT

Fant, G.: "Descriptive Analysis of the Acoustic Aspects of Speech", invited paper presented at the Wenner-Gren Foundation for Anthropological Research Symposium on Comparative Aspects of Human Communication at Burg Wartenstein/Austria, September 1960; to be publ. in LOGOS the Bulletin of the National Hospital for Speech Research

PUBLICATIONS AND REPORTS [x)]

Fant, G.: "Acoustic Theory of Speech Production", Mouton & Co., 's-Gravenhage 1960, 323 pp.

Fant, G.: "The Acoustics of Speech", invited paper presented at the 3rd International Congress on Acoustics, Stuttgart September 1959, to be publ. by Elsevier Publ. Co., Amsterdam, in the Proceedings from this congress

Fant, G.: "Descriptive Analysis of the Acoustic Aspects of Speech", invited paper presented at the Wenner-Gren Foundation for Anthropological Research Symposium on Comparative Aspects of Human Communication at Burg Wartenstein/Austria September 1960, to be publ. in LOGOS the Bulletin of the National Hospital for Speech Research

Fant, G., Stevens, K.N.: "Systems for Speech Compression", Fortschritte der Hochfrequenztechnik Vol. 5 (1960) 229-262, Akademische Verlagsgesellschaft M.B.H., Frankfurt am Main

Johansson, B.: "A New Coding Amplifier System for the Severely Hard of Hearing", paper presented at the 3rd International Congress on Acoustics, Stuttgart, Sept. 1959, to be publ. by Elsevier Publ. Co., Amsterdam, in the Proc. from this congress

Møller, A.: "Improved Technique for Detailed Measurements of the Middle Ear Impedance", J. Acoust.Soc.Am. 32, 250-257 (1960)

Møller, A.: "Network Model of the Middle Ear", to be publ. in J.Acoust.Soc.Am. 33, Febr. 1961

Møller, A.: "The Acoustic Impedance of the Human Ear", paper presented at the 3rd International Congress on Acoustics, Stuttgart, Sept. 1959, to be publ. in the Proc. from this congress

Wedenberg, E.: "Auditory Training of the Severely Hard of Hearing Using Coding Amplifier", paper presented at the 3rd International Congress on Acoustics, Stuttgart, Sept. 1959, to be publ. in the Proc. from this congress

[x)] All publications and reports which are under preparation or which have recently been published are included in this list.

STL-QPSR-1/1960 (July-September)

Speech Transmission Laboratory

QUARTERLY PROGRESS AND STATUS REPORT

October 15, 1960

SPEECH TRANSMISSION LABORATORY
Div. of Telegraphy-Telephony
Royal Institute of Technology
Stockholm/Sweden

STL-QPSR-1/1960 (July-September)

Speech Transmission Laboratory

QUARTERLY PROGRESS AND STATUS REPORT

October 15, 1960

SPEECH TRANSMISSION LABORATORY
Div. of Telegraphy-Telephony
Royal Institute of Technology
Stockholm/Sweden

ACKNOWLEDGMENTS

The research reported in this document was made possible in part by supports from the following sources:

		Chapter
State Council of Technical Research (Statens Tekniska Forskningsråd)	(1)	I,II,III,IV B,E
State Council of Medical Research (Statens Medicinska Forskningsråd)	(4)	I A, IV
Research Institute of the National Swedish Defence (Försvarets Forskningsanstalt)	(3)	I,II,III
Knut och Alice Wallenbergs Stiftelse	(4)	IV
other Swedish sources		IV,V
US Air Force under Contract AF 61(052)-342	(3)	I,II,III
US Army under Contract DA-91-591-EUC-1496 OI-4676-60	(2)	I

(1) Speech communication research on a broad basis
(2) Speech research, particularly speech analysis techniques
(3) Speech research, particularly speech analysis and synthesis
(4) Speech and hearing research of medical interest

TABLE OF CONTENTS

SPEECH TRANSMISSION LABORATORY x)

PERSONNEL

ADMINISTRATION AND PERMANENT STAFF

Speech Communication Research:

Tekn.dr.,Doc. G. Fant	Director of the Speech Transmission Laborator and the Speech Communication Research Group
Fru Marianne Richter	Secretary
Fru Si Felicetti	Secretary
Civ.ing. U. Rengman	Assistant manager and research associate
Ing. B. Wejnebring	Head of laboratory workshop
Civ.ing. A. Risberg	Assistant director of Speech Communication Research Group

Technical Audiology:

Ing. B. Johansson	Head of the Technical Audiology Group
Fru Ester Lindberg	Secretary

Hearing Research:

Dipl.ing. G. Rösler	(Karolinska Institutet, Fysiologisk Akustik)
Herr E. Voolahe	Audiometry technician

RESEARCH ASSOCIATES

Med.kand.,Civ.ing. C. Cederlund	fellowship from State Council of Technical Research
Civ.ing. J. Liljencrantz	
Fil.mag. B. Lindblom	
Dipl.ing.(E.T.H.) J. Mártony	
Ing. A. Møller	
Fil.dr. H.M. Truby	
Fil.kand. S. Öhman	

GUEST RESEARCHES AND TEMPORARY ASSOCIATES

Dr. E.C. Carterette	(Assistant professor of psychology, Universit; of California, Los Angeles/USA) National Science Foundation Postdoctoral Fell
Mr. H. Fujisaki	(Massachusetts Institute of Technology, Cambridge/USA)
Siv.ing. A. Krokstad	(Norges tekniske högskole, Trondheim/Norge) fellowship from Scandinavian Council for Applied Research
Siv.ing. M. Kringlebotn	(Norges tekniske högskole, Trondheim/Norge) fellowship from Scandinavian Council for Applied Research
Dipl.ing. W. Kurtze	
Dr. L. Lisker	(Haskins Laboratories, New York/USA)

x) The Speech Transmission Laboratory is a research department within the Division of Telegraphy-Telephony (Head Professor Torbern Laurent)

TECHNICAL ASSISTANTS

Herr L. Andersson
Herr S.-E. Appelgren x)
Herr S. Berg
Herr B. Lindström
Fru Ingrid Møller

THESIS STUDENTS

Teknolog G. Garpendahl
Teknolog J. Hedman
Teknolog H. Laul

PERSONNEL FROM INSTITUTIONS OUTSIDE THE R.I.T. CONTRIBUTING TO THE WORK MENTIONED IN THIS REPORT

Med.lic. G. Bjuggren	Sabbatsbergs Sjukhus, Stockholm
Med.lic. L. Björk	Akademiska Sjukhuset, Uppsala
Med.dr., Doc. T. Lundborg	Södersjukhuset, Stockholm
Med.lic. B. Nylén	Akademiska Sjukhuset, Uppsala
Med.lic. B. Nyström	Karolinska Sjukhuset, Stockholm
Med.dr., Doc. E. Wedenberg	Karolinska Institutet, Stockholm

x) Technical Audiology Group

PUBLICATIONS AND REPORTS

		status:
Fant, G.:	"Acoustic Theory of Speech Production" to be publ. by Mouton & Co., s'-Gravenhage	2nd proof
Fant, G.:	"The Acoustics of Speech" invited paper presented at the 3rd International Congress on Acoustics, Stuttgart September 1959, to be publ. by Elsevier Publ.Co., Amsterdam, in the Proceedings from this congress	
Fant, G.:	"Descriptive Analysis of the Acoustic Aspects of Speech" invited paper presented at the Wenner-Gren Foundation for Anthropological Research Symposium on Comparative Aspects of Human Communication at Burg Wartenstein/Austria September 1960, to be publ. in LOGOS the Bulletin of the National Hospital for Speech Research	
Fant, G., Stevens, K.N.:	"Systems for Speech Compression" to be publ. by Fr. Rühmann, Karlsruhe-Durlach, in FORTSCHRITTE DER HOCHFREQUENZTECHNIK	2nd proof, preprint
Johnasson, B.:	"A New Coding Amplifier System for the Severely Hard of Hearing" paper presented at the 3rd International Congress on Acoustics, Stuttgart September 1959, to be publ. by Elsevier Publ.Co., Amsterdam, in the Proceedings from this congress	
Møller, A.:	"A Network Model of the Middle Ear" R.I.T. Stockholm, Sweden, Report No. 17, June 8, 1960, Speech Transmission Laboratory, to be publ. in revised form in J.Acoust.Soc.Am.	
Møller, A.:	"The Acoustic Impedance of the Human Ear" paper presented at the 3rd International Congress on Acoustics, Stuttgart September 1959, to be publ. by Elsevier Publ.Co., Amsterdam, in the Proceedings from this congress	
Wedenberg, E.:	"Auditory Training of the Severely Hard of Hearing Using Coding Amplifier" paper presented at the 3rd International Congress on Acoustics, Stuttgart September 1959, to be publ. by Elsevier Publ.Co., Amsterdam, in the Proceedings from this congress.	

INTRODUCTION

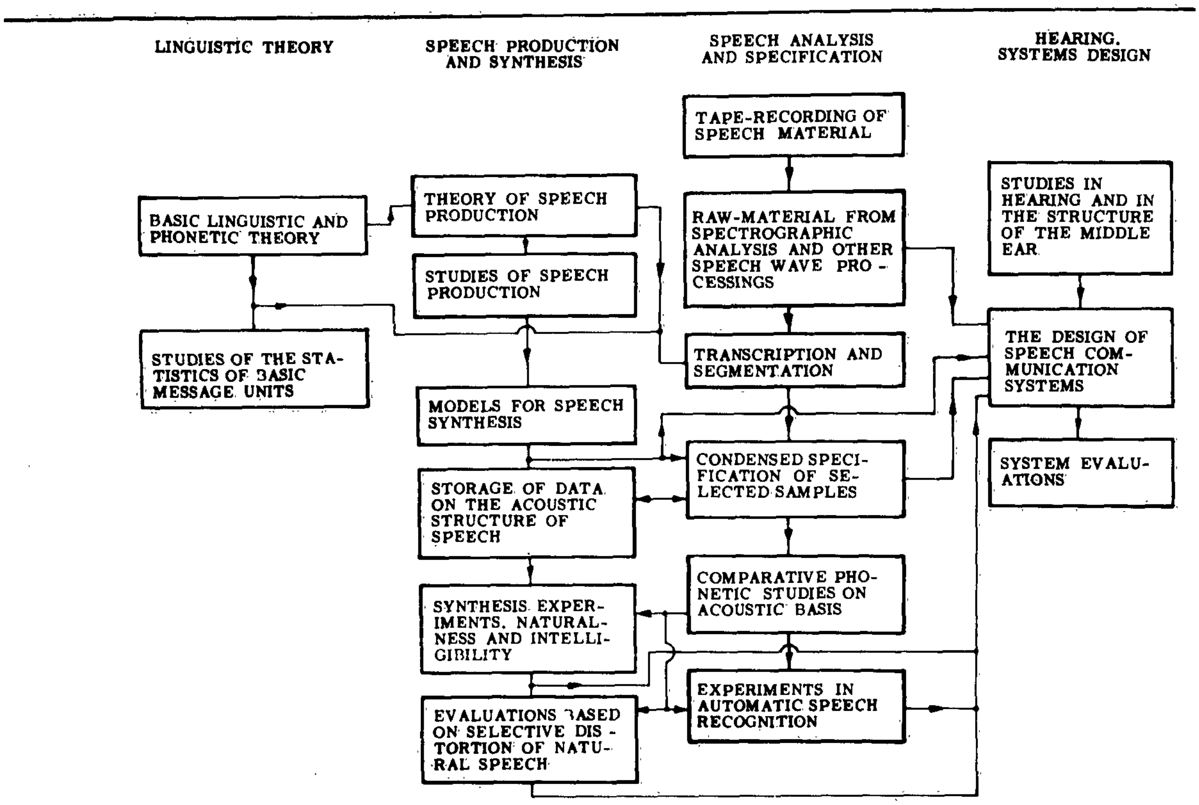
This report series is intended for quarterly summaries of recent progress and status of research at the Speech Transmission Laboratory of the Royal Institute of Technology, Stockholm. The present issue is the first report of the series intended for general distribution and a comparatively large portion of the contents is therefore devoted to a presentation of the general scope and present standing of active projects. Any particular results reviewed here are to be considered as preliminary only and will normally reappear in coming scientific papers.

Stockholm, October 15, 1960

Gunnar Fant

PLAN OF SPEECH COMMUNICATION RESEARCH

PRESENT AND PROPOSED ACTIVITIES


LINGUISTIC THEORY
SPEECH PRODUCTION AND SYNTHESIS
SPEECH ANALYSIS AND SPECIFICATION
HEARING. SYSTEMS DESIGN
TAPE-RECORDING OF SPEECH MATERIAL
BASIC LINGUISTIC AND PHONETIC THEORY
THEORY OF SPEECH PRODUCTION
RAW-MATERIAL FROM SPECTROGRAPHIC ANALYSIS AND OTHER SPEECH WAVE PRO - CESSINGS
STUDIES IN HEARING AND IN THE STRUCTURE OF THE MIDDLE EAR
STUDIES OF SPEECH PRODUCTION
STUDIES OF THE STATISTICS OF BASIC MESSAGE UNITS
TRANSCRIPTION AND SEGMENTATION
THE DESIGN OF SPEECH COMMUNICATION SYSTEMS
MODELS FOR SPEECH SYNTHESIS
CONDENSED SPECIFICATION OF SELECTED SAMPLES
SYSTEM EVALUATIONS
STORAGE OF DATA ON THE ACOUSTIC STRUCTURE OF SPEECH
COMPARATIVE PHONETIC STUDIES ON ACOUSTIC BASIS
SYNTHESIS EXPERIMENTS. NATURALNESS AND INTELLIGIBILITY
EXPERIMENTS IN AUTOMATIC SPEECH RECOGNITION
EVALUATIONS BASED ON SELECTIVE DISTORTION OF NATURAL SPEECH

RESEARCH OBJECTIVES

KNOWLEDGE OF THE STATISTICS OF SPEECH MESSAGE SIGNS

EXPERIENCE OF VARIOUS METHODS OF SPEECH ANALYSIS AND SYNTHESIS

KNOWLEDGE OF THE AUDITORY FUNCTIONS AND OF THE MIDDLE EAR STRUCTURE

KNOWLEDGE OF THE PHYSICAL STRUCTURE OF SPEECH, PRIMARILY ON THE LEVELS OF SPEECH PRODUCTION AND THE ACOUSTIC SPEECH WAVE

RECOMMENDATIONS FOR THE DESIGN OF MAXIMALLY EFFICIENT SPEECH COMMUNICATION SYSTEMS.
SPEECH COMPRESSION SYSTEMS.
AIDS FOR THE DEAF AND THE BLIND

THEORY AND RULES FOR TRANSLATING FROM SPEECH PRODUCTION (ARTICULATION) DATA TO SPEECH WAVE DATA AND VICE VERSA

LANGUAGE DESCRIPTIONS AND PHONETIC SYSTEMATIZATION ON ACOUSTIC BASIS

RULES FOR SPEECH SYNTHESIS FROM MESSAGE SIGNS

RULES FOR MACHINE RECOGNITION OF SPEECH

I. SPEECH ANALYSIS

A. VOICE SOURCE STUDIES

The inverse filtering techniques of studying the source of voiced sounds which have been reported on earlier provide data of the type shown in Fig. I-1. The speech wave is passed through a filter network comprising four variable anti-resonance circuits and a circuit representing the inverse of a higher pole correction. The present studies have been based on sung vowels and the investigator has adjusted the anti-resonance frequencies so as to attain a maximal cancellation of the formant ripple.

A correct shape of the glottal flow pulses is obtained only under the conditions of negligible phase distortion down to 40 c/s. A normal laboratory tape-recorder does not meet these requirements and produces a distinct wave-form distortion. A high frequency modulation circuit was utilized in conjunction with the condenser microphone in order to preserve a response down to DC. It was found that instabilities in the atmospheric pressure gave rise to very low disturbances which tend to throw the glottal pulse train picture out of the range of the oscillograph. These effects have been reduced by the use of a phase compensated high-pass filter of 20 c/s cutoff frequency. An FM-modulation system (1) is under construction and will be used in future experiments for the recording of connected speech material.

The following results have been obtained so far:

(1) The shape of a glottal flow pulse in the chest register varies from a smooth almost sinusoidal shape at low voice efforts to an approximate triangular shape at high voice efforts. The closed interval within a complete period of glottal opening and closing is of relatively short duration at very low voice efforts and occupies up to 75 % of the period at very high voice efforts.

(2) The peak amplitude of the glottal volume velocity pulse changes much less than the overall amplitude of the sung vowel. This observation correlates well with the relative stability within connected speech of the intensity of the voice fundamental. Experiments on 5 subjects showed that the peak amplitude of the glottal flow increased on the average 4 dB per 10 dB increase of the overall intensity of the vowel. These figures are

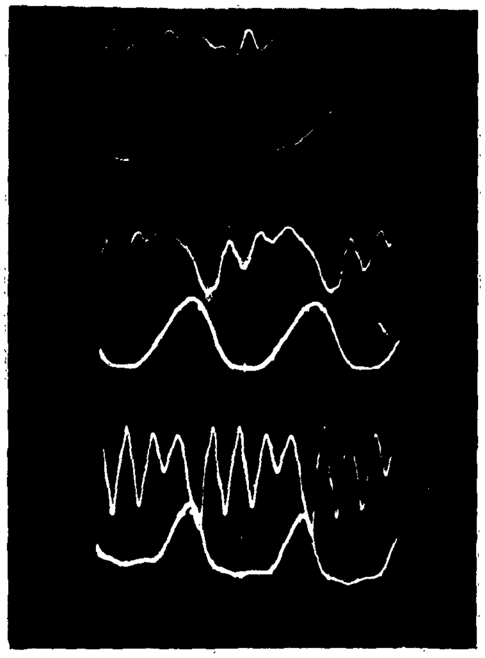


Fig. I-1 Oscillograms of the vowel [æ:] and of the regenerated glottal flow function. Low, medium, and high voice efforts are illustrated from the top to the bottom of the figure.

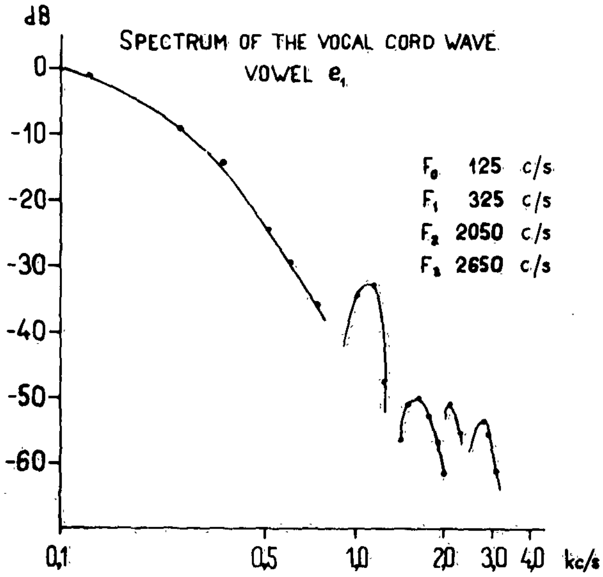


Fig. I-2 Source spectrum of a sung vowel. The spectral minimum at 700 c/s is a typical feature and is the frequency domain correspondence of a vocal pulse base length of 2/700 = 0.0028 sec.

the same as those reported for the relative variations of the amplitude of the voice fundamental in sung vowels. (2) It remains to be seen to what extent this constancy appears in connected speech.

(3) Sufficiently representative data on the spectra of the regenerated glottal flow function at various voice levels have not yet been collected but it is apparent from the discussion above as well as from other investigations, e.g., Miller (3), that high voice level implies higher efficiency and a relative boost of the upper part of the spectrum. The following source spectrum may exemplify a medium-high voice effort (Fig. I-2).

(4) Confirming earlier observations, e.g., Miller (3), it was found that the apparent starting point of the formant oscillation within a voice cycle is typically close to the instant of closing of the vocal cords. During the open glottis interval the formant oscillation is often markedly reduced in amplitude. The coupling to the subglottal system thus causes appreciable damping. As could be expected owing to the flow dependent resistance this damping is larger at low voice efforts than at high voice efforts. The work is continuing with theoretical studies of the time domain characteristics of vowels from Laplace transforms assuming excitation functions of various shapes.

C. Cederlund, A. Krokstad, M. Kringlebotn

(1) Garpendahl, G.: "Frekvensmodulator för lågpassregistrering på bandspelare", thesis work under progress (1960).

(2) Fant, G.: "Acoustic analysis and synthesis of speech with applications to Swedish", Ericsson Technics, 15, No. 1, 3-108 (1959).

(3) Miller, R.L.: "Nature of the vocal cord wave", J.Acoust.Soc.Am. 31, 667-677 (1959).

B. VOICE FUNDAMENTAL FREQUENCY TRACKING

Several systems for F_0-tracking have been tried in the past with an effort to construct a relatively simple instrumentation for use in vocoders and for phonetic research. Our experience supports the general view that any system will work fine on some voices and particularly well in sustained portions of speech or singing. No simple system, however, has been considered reliable enough for vocoder usage and all systems have had the weakness of being sensitive to hum, noise, and statics from the voice channel. The most common error remaining, in the case of a high quality speech input, is the tendency towards synchronization on harmonics or the temporary indication of a subharmonic. We have tried various non-linear systems for regenerating or enhancing the voice fundamental and in combination with the following prefiltering:

(1) A fixed low-pass filter optimally selected for the particular speaker.

(2) A low-pass filter or a band-pass filter continuously controlled by the measured output of the F_0-meter.

(3) Three band-pass filters combined with logics for selecting as the input to the frequency-measuring stage the output of the band-pass filter of lowest center frequency carrying signal above a certain threshold value.

None of these have functioned to our satisfaction. System number 1 is as good as any of the other two. System number 2 is subject to errors due to the delay in frequency measurements and in starting errors. System number 3 is sensitive to switching transients and to unfavorable phase combinations of signals from two band-pass channels.

A substantial gain in accuracy has recently been obtained in an experimental set up which is similar to system number 3 above in some respects. The basic idea is to avoid time-variable filtering and to incorporate one complete frequency counter in each of the band-pass channels and to decide which channel provides the lowest frequency measure. This measure is selected as the most probable F_0. Errors due to synchronization on overtones are avoided provided one of the channels carries a sufficiently pure fundamental. Successful results have been reached with a two-channel system. A three-channel system comprising three complete F_0-meters and a min-

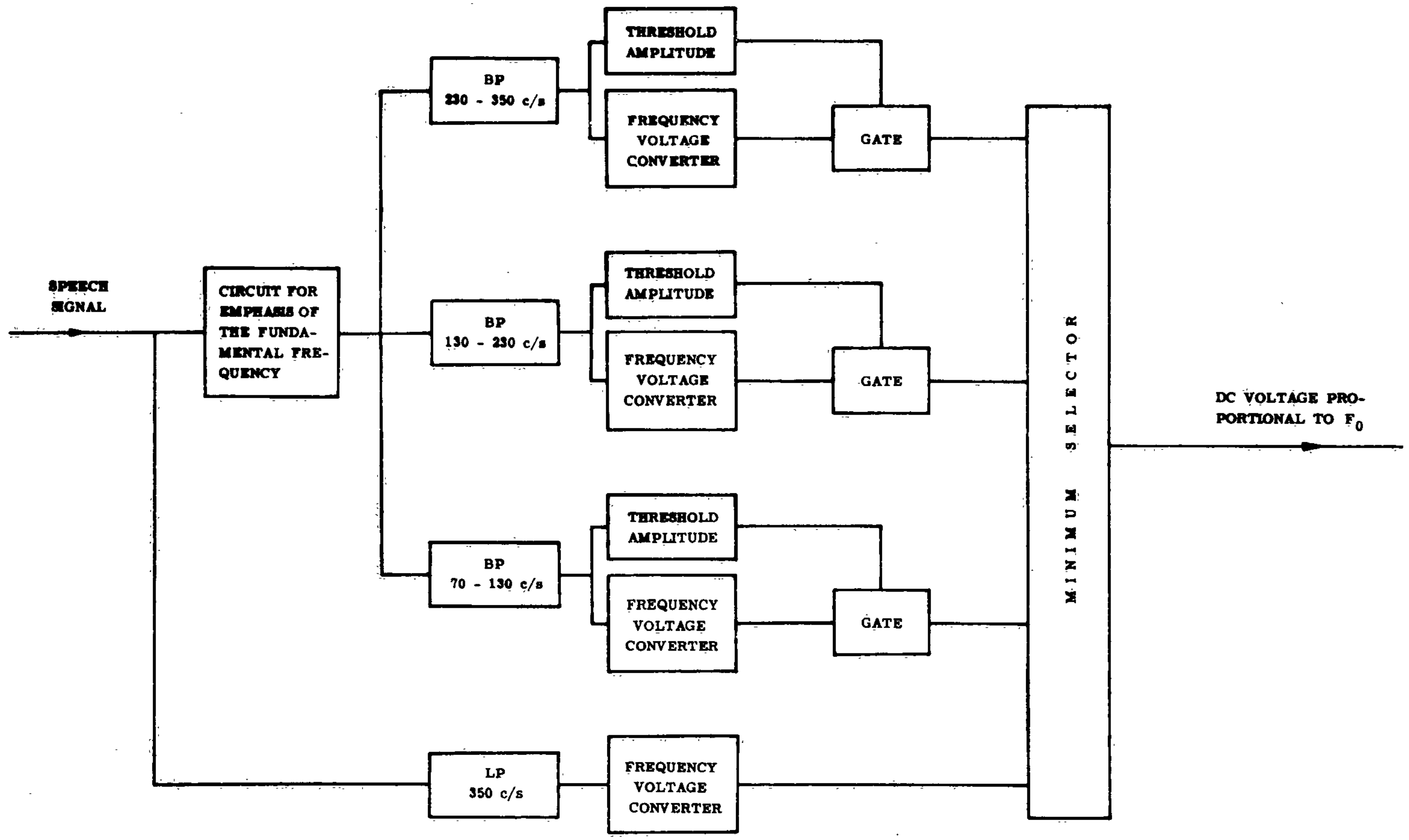

SPEECH SIGNAL
CIRCUIT FOR EMPHASIS OF THE FUNDAMENTAL FREQUENCY
BP 230 - 350 c/s
BP 130 - 230 c/s
BP 70 - 130 c/s
LP 350 c/s
THRESHOLD AMPLITUDE
FREQUENCY VOLTAGE CONVERTER
THRESHOLD AMPLITUDE
FREQUENCY VOLTAGE CONVERTER
THRESHOLD AMPLITUDE
FREQUENCY VOLTAGE CONVERTER
FREQUENCY VOLTAGE CONVERTER
GATE
GATE
GATE
MINIMUM SELECTOR
DC VOLTAGE PROPORTIONAL TO F_0

Fig. I-3 Proposed scheme for F_0-tracking.

imum selector is under construction. The basic system is illustrated in Fig. I-3.

The success of this system, or of any other frequency-measuring system, will depend on the actual presence of a voice fundamental of an amplitude which may not be much less than that of any harmonic. Special attention has therefore been devoted to the initial stage for enhancing the voice fundamental. A few simple non-linear systems have been tested recently. A speech material of 10 seconds each from 4 male and 4 female speakers was processed by the various methods. Narrow-band spectrograms of the results were studied and evaluated with regard to the relative intensity of the voice fundamental. The percentage of pitch periods which were judged to require only a moderate filtering in the following stages were counted. The following results were obtained:

Voice channel 50-3000 c/s input

Method / Speaker	Direct	Half-wave rectification		Full-wave rectification	Rectified single side band
		phase 1	phase 2		
Male	49	35	50	47	83
Female	85	67	87	22	82

These results do not pertain to the overall performance of a complete F_0-meter. The half-wave rectification is phase sensitive. The direct unprocessed speech provides the best material for female voices which is due to the natural prominence of their fundamental. Full-wave rectification tends to produce a frequency multiplication. In these instances the second harmonic is highly boosted which accounts for the low figure of merit, 22 % for female voices.

Rectified single side band provides the best results. The shortcomings of the single side-band processing are mostly due to instances in which the speech wave either was of low intensity or was dominated by the voice fundamental. These findings conform with results from a theoretical analysis made by H. Fujisaki. (1) At low signal levels the rectifier characteristics approximated a square-law function which accounts for a second power dependency of the amplitude of modulation products on the input signal amplitude. In case the input signal has a flat spectrum envelope and the harmonics are linearly related in phase it may be shown that the ratio of the fundamental to the second harmonic after the SSB-rectification, becomes (N-1)/(N-2), where N is the number of harmonics present in the input. An input band consisting of two harmonics is thus optimal and it has been shown that the presence of a formant structure will favorably influence the ratio of fundamental to second harmonic.

A separate low-pass channel plus frequency counter incorporated in a pitch meter should favorably supplement the part of the system fed from a single side-band input. Other alternatives exist in parallel systems based on different selection of the input voice band, e.g., high-pass filtering before the SSB-operation. Inverse-filtering methods might produce a raw material for frequency counting competing with the SSB-methods. A requirement is then that the base band in the F_0-region shall be intact. Preliminary studies have given promising results but further studies are needed.

A. Risberg, A. Møller, H. Fujisaki

(1) Fujisaki, H.: "Theoretical studies on pitch extension and formant tracking", internal STL-report, Aug. 20 (1960).

C. FORMANT FREQUENCY TRACKING

The scope of this project is to gain experience on various schemes of formant tracking, i.e., automatic extraction of formant frequency data from connected speech.

During the past year we have made some preliminary studies on methods of counting zero-crossings and of moment weighting. The moment-weighting method was engineered simply by carrying out the analog division between the differentiated speech wave and the undifferentiated speech wave.

The recent work during the last three months has been concerned with more detailed studies of zero-counting techniques and with a method of anti-resonance filtering. When the zero of an anti-resonance circuit coincides in frequency and bandwidth with a formant pole there is perfect cancellation providing the residues from other poles may be neglected. This procedure may be regarded as the time-domain equivalent of the pole-zero spectral-matching method developed at M.I.T.

Our results are negative in so far as all the methods mentioned above have serious inherent limitations, at least within the experimental conditions of our present studies. Our main conclusions are as follows:

(1) Zero-crossing counting.

a. If the frequency of the damped oscillation corresponding to a single formant is integrated over a period longer than that of a voice fundamental period the measure will coincide with the frequency of the most intense harmonic within the formant. This well-known effect will cause objectionable jitter effects in resonance vocoders unless an excessively long time constant is chosen for the frequency parameter smoothing.

b. An initial transposition of the speech band to higher frequencies will make possible the use of smoothing time constants much smaller than that of a voice fundamental period. However, the presence of the discontinuity at the time position of the initial voice excitation and also the frequent absence of oscillation in a part of the cycle will complicate the measuring procedure by the need of an intricate sampling system.

c. The presence of insufficiently suppressed residues from other formants might distort the measurement owing to averaging effects. If two frequency components are less than one octave apart this effect becomes negligible as long as the unwanted signal is more than 5 dB below the level of the signal to be measured. As the frequency ratio increases the effect is more pronounced. In the case of a frequency ratio of $F_2/F_1 = 10$ and an amplitude ratio of $A_2/A_1 = 10$ dB there results an error of $0.15(F_2-F_1)$. A prefiltering continuously controlled by the measured positions of other formants is needed (1) in order to get the best results from the method. A supplementary method which has been tried is to restrict the time constant of the flipflop in the frequency counter circuit so as to avoid synchronization on interfering oscillations from high-frequency formants.

(2) Moment weighting.

a. Moment-weighting and zero-crossing counting techniques are known to provide equal results on random noise signals. In general both methods are dependent on a prefiltering.

b. Our experiments on the use of the simple moment-weighting technique of analog division of two speech signals differing by a prefiltering of 6 dB/octave and subjected to linear rectification and short-time smoothing, have shown that the function is very sensitive to the particular prefiltering and to the degree of asymmetry of the formant. In case of reasonably symmetric formants the systems tend to follow the leading harmonic.

c. The moment-weighting procedure may from a systematic point of view be regarded as a simplified instance of interpolation within a band of filters connected to a maximum selector. Such systems have been tried with some success in England (2) and will be evaluated in coming phases of our work.

(3) Anti-resonance filtering.

A pilot study on anti-resonance methods of formant tracking has recently been completed. The ideal system would be one in which the speech wave is passed through a series of anti-resonance circuits in cascade, one for each formant. By proper adjustment of

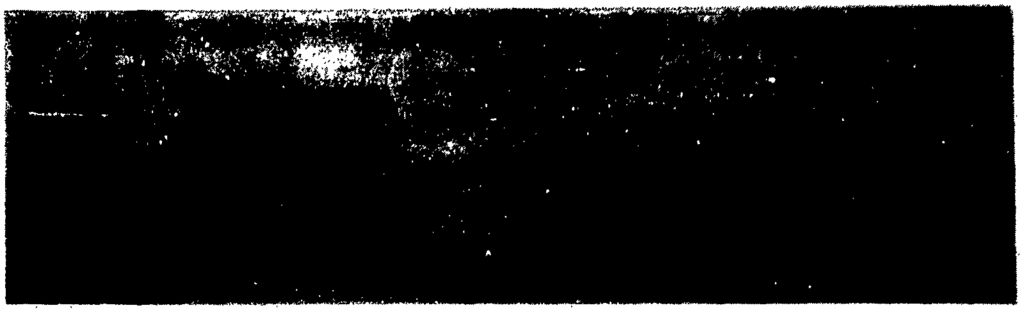


Fig. I-4 Spectrogram with superimposed F_1 - values originating from a minimum selection of the outputs of a anti-resonance curcuit spaced with 100 c/s intervals.

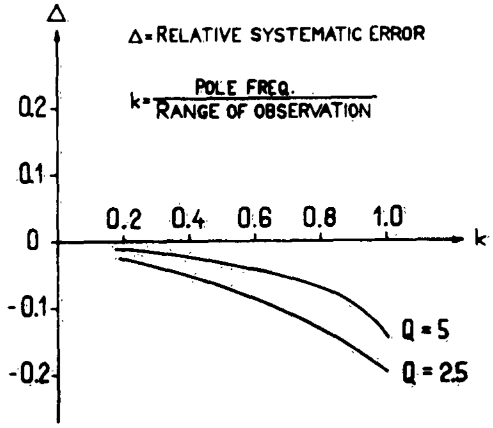


Fig. I-5 Systematic error in the tracking of a one-formant vowel owing to prefiltering with a LP filter.

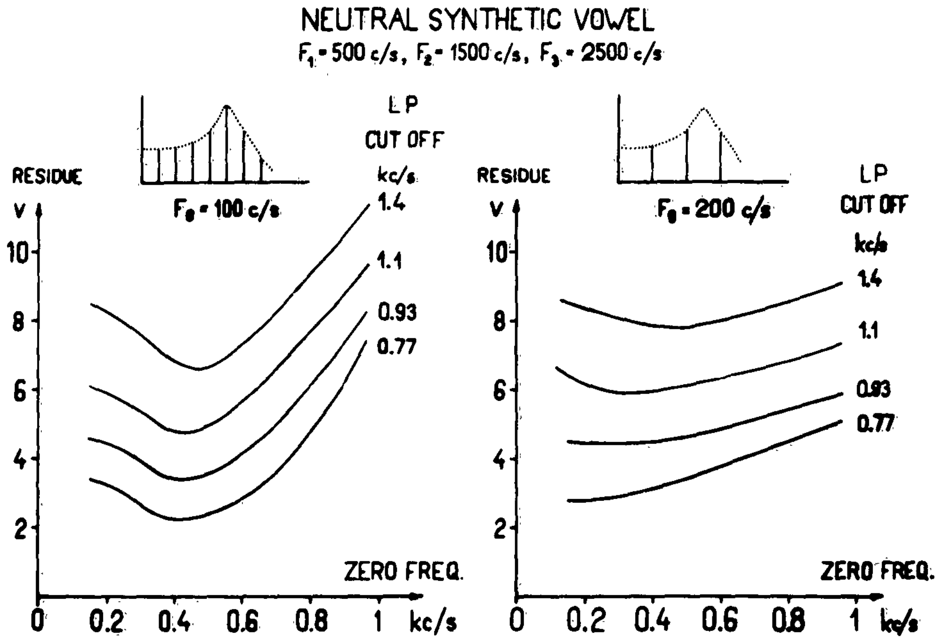


Fig. I-6 Anti-resonance analysis of the F_1 - range of a neutral vowel at two different F_0 - pitches and varying prefiltering LP cut off frequency.

the frequency and bandwidth of each anti-resonance it would be possible to cancel the damped oscillations of all formants and there would remain merely a voice excitation function. This is actually the method utilized in the voice-source studies, section I-A of this report. In formant-tracking systems, however, it is necessary to adopt a matching criterion that is independent of a human operator.

Our efforts have concentrated on the theoretical and experimental study of tracking the first formant after it has been isolated by a prefiltering stage. The theoretical approach was to calculate the intensity output of the anti-resonance circuit as a continuous function of its center frequency setting. The bandwidth of the zero was assumed to coincide with that of the formant pole but an exact match is not critical. The experimental set up consisted of measuring the linear rectified and smoothed output of an anti-resonance circuit varied in steps of 100 c/s between successive recordings thus simulating a bank of parallel channels of anti-resonance filters. The F_1-frequency was selected as the channel carrying minimum intensity. A fairly promising result was obtained in a test with a piece of connected human speech as can be seen in Fig. I-4, where the data are superimposed upon a spectrogram. A theoretical analysis of the anti-resonance method provides the following results:

a. There is a systematic error when the low-pass prefiltering range is narrowed as may be seen from Fig. I-5. This error is approximately inversely proportional to the Q of the formant and amounts to 10 % when the cutoff of the prefilter is 10 % above the formant frequency assuming a Q of 5 and further a very low F_0 and a negligible influence from higher formants.

b. Systematic errors of greater importance occur when the voice fundamental frequency, F_0, is high and especially when the formant peak falls halfway between two harmonics. This is illustrated in Fig. I-6, which pertains to the experimental analysis of a standard neutral vowel [ɜ] (F_1=500, F_2=1500, F_3=2500 c/s).

At the pitch of F_0=200 c/s the selectivity is very bad and a minimum is obtained only with the widest possible prefilter.

The fundamental weakness of the approach reported on above is that only a limited part of the spectrum is allowed to contribute to the minimum response whereas a frequency shift of a formant affects the entire spectrum. The sensitivity of the method is upset by incomplete cancellation of residues from other formants and from the vocal-cord wave. A cascaded system of several anti-resonances, one for each formant, would do better but is more complicated for automatic tracking systems. The supplementary usage of phase information might provide greater accuracy.*

H. Fujisaki, A. Risberg

* Such a system is that of Lawrence (personal communication) who works with a system of two self-adjusting anti-resonance circuits which continuously follow the first two formants of speech. His criterion for self-adjustment is the relative phase between the output and the input of the anti-resonance circuit.

(1) Chang, S.-H.: "Two schemes of speech compression system", J.Acoust. Soc.Am. 28, 565-572 (1956).

(2) Holmes, J.N.: "A method of tracking formants which remains effective in the frequency regions common to two formants", Res. Rep. JU8-2, Joint Speech Research Unit, B.P.O., Dec. 1958; and

Holmes, J.N., Kelly, L.C.: "Apparatus for segmenting the formant frequency regions of a speech signal", Res. Rep. JU9-4, Joint Speech Research Unit, B.P.O., Aug. 1959.

D. AUTOMATIC IDENTIFICATION OF SOUND FEATURES

1. Detection of voicing

Several systems for an automatic detection of voicing have been looked into. Methods based on the relative distribution of spectral energy as determined from various combinations of low-pass and high-pass filters have proved to be less successful than methods which rely on the intensity in a low frequency band compared with the average value of the syllabic intensity of the subject's speech. The accuracy of voicing detection may be further improved by the linear subtraction of a running measure from a zero-counting circuit operating on the unfiltered input speech. Voiced sounds have lower zero-crossing rates than unvoiced sounds.

2. Automatic segmentation schemes

An inspection of spectrograms reveals the existence of natural acoustic boundaries between speech sounds or rather speech segments which may constitute a part of a speech sound. The criterion for the presence of such a boundary may be defined from the rate of change of spectral energy with respect to time in several band-pass limited regions of the spectrum. A linear summation of the absolute values of the rates of spectral changes provides a sensitive measure of the degree of spectral discontinuity. An experimental arrangement consisting of band-pass filters, rectifiers, smoothing filters, differentiating circuits, rectifiers, and a final smoothing and connection to a summation stage for several parallel channels has been constructed. Preliminary experiments have provided promising results.

J. Liljencrantz

E. EVALUATION OF SPECTROGRAPHIC DATA SAMPLING TECHNIQUES

1. Sectioning of unvoiced sounds with the Sonagraph

One weakness of the sectioning device on the Sonagraph spectrum analyzer is the short-time constant in the smoothing RC-filter following the rectifier circuit. This time constant, calculated with the correction for the speed-up factor in playback, is of the order of 8 msec which is rather short for a correct wide band (300 c/s) sectioning of random noise sounds such as fricatives and stops. As a general rule the averaging time of the smoothing filter should be much greater than the averaging time of the band-pass filter which implies a cutoff frequency of the smoothing filter much smaller than the bandwidth of the analyzing filter (1). If this requirement is not met with it may be expected that the random fluctuations superimposed on the spectral section will be large, i.e., there will appear a fine structure in the spectral section which is due to the general statistical properties of filtered random noise alone. The correspondence in the time-frequency-intensity spectrogram is the presence of the random striations. These have continuity in the frequency domain in the form of peaks and valleys which in a section may be confused with formants and anti-resonance effects. The approximate extent in time and frequency of a striation maximum is governed by the law of reciprocal spread. The expected average spacing of random maxima in the frequency domain as well as the expected number of envelope maxima per second in the time domain should be of the same order of magnitude as the bandwidth of the band-pass filter, or of a formant within the analysis filter whichever is the smaller and thus has the greater inertia effect.

The uncertainty of any ordinate within a spectral section of a random noise sound expressed as the ratio of root mean square random error to the expected long-time average value is proportional to $(B_s/B_a)^{\frac{1}{2}}$ where B_s is the width of the smoothing filter and B_a is the width of the analysis filter. (1)(2)

Accordingly, the time constant of the integration circuit in the Sonagraph sectioner was increased to 60 msec effective time which is well above that of the broad-band analysis filter. This change has markedly improved the reliability of spectral sections or fricatives taken with the Sonagraph but there remains a certain mechanical instability in the function of the microswitch which introduces an uncertainty in the exact location of a section of the order of 20 msec. Furthermore the dynamic range of the Sonagraph is not very large (30 dB) and one must be careful so as to avoid intermodulation effects.

For taking spectral sections of unvoiced sounds we generally rely on other instrumentations than the Sonagraph, e.g., repeated oscillographic recordings of the output of a wave analyzer. However, for less critical applications the Sonagraph sectioner is of some use.

2. Formant frequency measurements with the Sonagraph

An experimental study has been undertaken of the consistency in a subject's repeated measurements of the center frequencies of the first four formants from wide-band time-frequency-intensity spectrograms of human vowels (F_0 of the order of 120 c/s) and the spread among five subjects owing to individual variations in the rationale for determining formant center frequencies. The average deviation from the mean of repeated measurements by a single subject ranged from 0.2 mm t● 0.6 mm corresponding to a spread of 10-30 c/s. These figures pertain to mean values for four formants measured in six vowels. The inconsistency of a single subject's measurements varies inversely with the distance from a formant to the nearest formant in a spectrum and is maximally of the order of 150 c/s but stray values up to 300 c/s were found.

The systematic disagreement between subjects was maximally 130 c/s with an average value of the order of 50 c/s. This systematic spread is slightly higher than that reported by Flanagan.[3]

A prefiltering in the form of a 6 dB/octave base attenuation in the F1-region caused a shift up in the estimated position of F_1 by 40 c/s for natural speech but considerably less for synthetic speech. These figures pertain to male speakers and the effect is probably greater for female speakers with a higher F_0.

From measurements on synthetic vowels it was found that even in a broad-band spectrogram of vowels produced at a low F_0 of the order of 100-200 c/s there is a slight tendency to locate the center of the formant away from the ideal pole frequency value in a direction which is dependent on the particular F_0 and F_1 and basically the fine structure of the formant in a narrow-band section. At pitch values above $F_0 = 200$ c/s a less experienced investigator will tend to select a harmonic instead of interpolating between the harmonics, and the size of the maximum error thus approaches $F_0/2$.

A comparison study of section versus spectrogram is under way.

B. Lindblom, S. Öhman, A. Risberg

(1) Morrow, C.T.: "Averaging time and data-reduction time for random vibration spectra", Part I in J.Acoust.Soc.Am. <u>30</u>, 456-461 (1958); and Part II in J.Acoust.Soc.Am. <u>30</u>, 572-578 (1958).

(2) Beranek, L.: <u>Acoustic Measurements</u> (New York, 1949).

(3) **Flanagan,** J.L.: "A speech analyzer for a formant-coding compression system", Scientific Report No. 4, U.S. Air Force Contract No. AF 19(604)-626, May 1955, M.I.T., Acoust. Lab.

F. POLE-ZERO MATCHING TECHNIQUES

Our work in this field is still in an exploratory phase. Vowels are studied by means of the anti-resonance filter techniques mentioned in section I A. Most of the pole-zero matching of fricatives was made on a graphical basis comparing the spectra of human samples with spectra synthesized numerically from tabulated data of elementary pole and zero curves. Analog methods based on the use of networks with variable poles and zeros are also being investigated. A standard circuit for representing a pole-zero pair has recently been developed (Kringlebotn). It is based on the continuous variation of an inductance by means of feedback amplifier techniques.

A match of fricatives in terms of two poles and one zero is generally sufficient for retaining a high standard of speech quality in a formant-coded synthesis (OVE II).

A matching of the fricatives [s] and [f] in terms of two poles and one zero is shown in Fig. I-7. The measured samples pertain to sustained sounds analyzed by a closed loop process with a wave analyzer of 125-c/s bandwidth. The pole at 2700 c/s and the zero at 2500 c/s of the fricative [f] constitute a bound pair with but small contribution to the spectrum. It is of interest to see that the spectrum level rises all the way up to 12 kc/s which was the upper limit of analysis. Spectra of [f] vary much owing to the particular coarticulation and to the degree of labiodental constriction.

The main peak of the [s]-spectrum of Fig. I-7 is associated with the pole at 5800 c/s. The second pole at 8000 c/s contributes to build up a proper spectrum level at higher frequencies. The zero at 4500 c/s is placed higher than the corresponding zero in the measured spectrum in order to preserve a correct level ratio between the main formant and the low frequency part of the spectrum.

An additional inventory of two pole-zero pairs were added for matching the [s]-spectrum of Fig. I-8. One of these bound poles, the one at 2500 c/s, corresponds to F3 and the one at 4200 c/s to F5. These weaker formants do not appear to be necessary for the synthesis of a good [s].

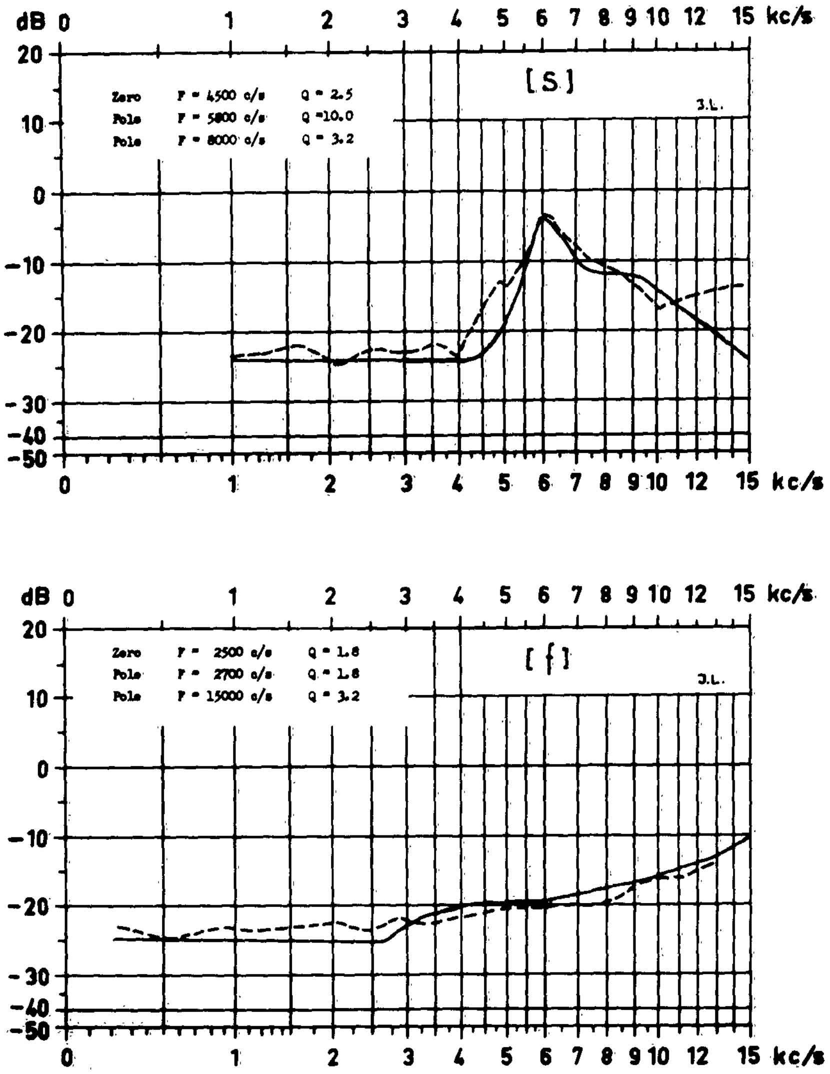


Fig. I-7 Measured spectra (broken lines) and two-pole-one-zero synthetic approximations (solid lines) of the fricatives [s] and [f].

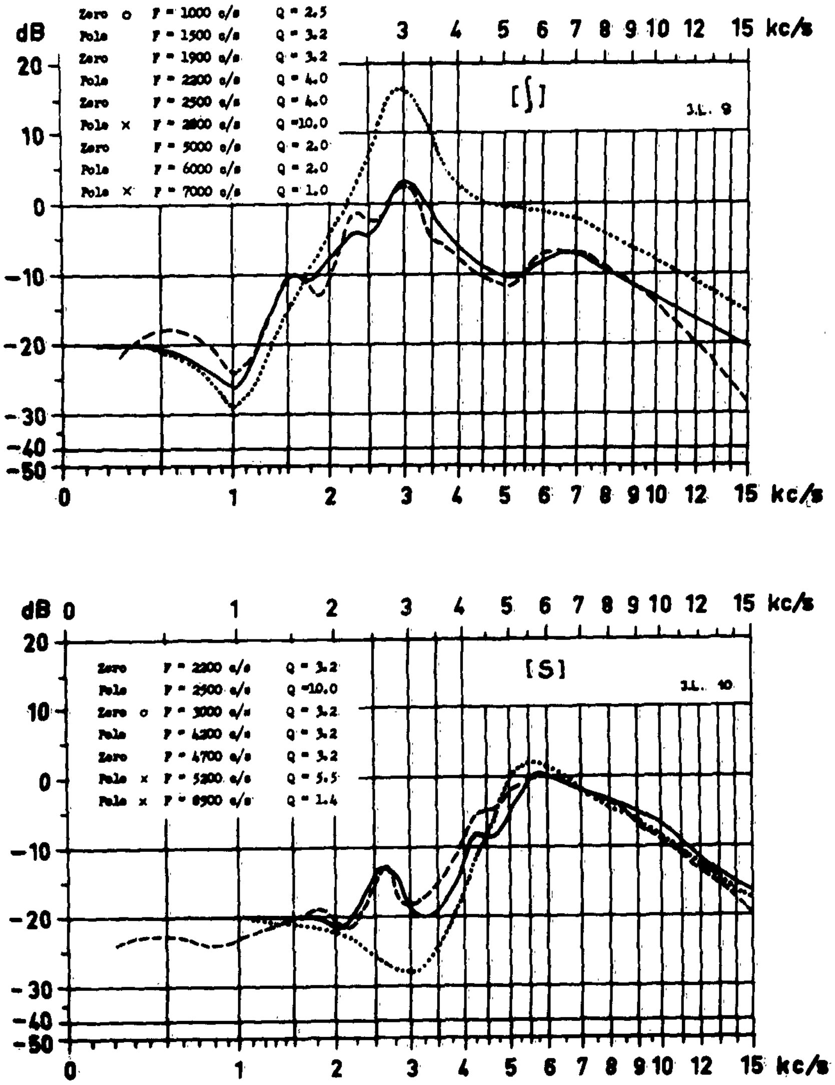


Fig. I-8 Pole-zero matching of [ʃ] and [s]. Measured spectra are represented by broken lines, two-pole-one-zero approximations by dotted lines and more complete synthetic representations comprising additional bound pole-zero pairs by solid line curves.
Free poles are marked × and free zeros are marked ○.

The spectrum of a fricative [ʃ] and its pole-zero approximation is demonstrated in the lower part of Fig. I-8. The essential feature of this particular palatal retroflex sound is a free zero at 1000 c/s and a free pole at 2800 c/s and one at 7000 c/s. A detailed match employing three additional pole-zero pairs associated with the relatively suppressed F2, F3, and F6 provides a match within a few dB from 300 c/s to 12 kc/s. The dotted curve on the figure pertains to the approximation in terms of the two free poles and the free zero alone. It is apparent that the resulting exaggeration of the relative level of the main peak is due to the absence of the high-frequency attenuation inherent in the two bound pole-zero pairs. This effect has been predicted in earlier theoretical work.[1] In agreement with results from those earlier studies [1] it is apparent that the synthesis can be made on the basis of a relatively flat source spectrum.

The pole-zero matchings performed in Fig. I-7 and I-8 allow a simplified structural comparison of the sounds [f], [s], and [ʃ]. There is a similarity between [s] and [ʃ] in so far as the spectra of both possess a free zero of a frequency lower than that of the two free poles. This free zero contributes effectively to the high-pass structure of the spectrum above the zero the significant part of which extends approximately 2000 c/s lower down in frequency for [ʃ] than for [s]. The mode spectrum of [f] does not possess a free zero and the only free pole is located at very high frequencies and is generally heavily damped. This pattern explains in part the relatively low overall intensity of [f].

An alternative interpretation applicable to the theory of distinctive features [2] would be to oppose [ʃ] to [s] and [f] as being the only sound that has a free zero below or close to F_2. This is a requirement for emphasizing formants F3 and F4 and also F2 if the zero is well below F_2 and thus a formant area in the consonant not far above the mean pitch of the upper formants of a following vowel. This conforms with the criterion of a major energy concentration in a centrally located peak as required by the definition of compactness. After correction of the [s]- and [f]-spectra for the relatively low sensitivity of the ear in the high-frequency region it is apparent that the [s]-spectrum has a higher center of gravity than the [f]-spectrum and [f] is thus grave compared with [s]. However, in some languages (e.g., in Swedish) it is feasible to oppose [ʃ] to [s] as being more flat (shift down of the spectrum) while other fricatives, e.g., [ç] take the place of the compact member of the system.

G. Fant, J. Mártony

(1) Fant, G.: Acoustic Theory of Speech Production ('s-Gravenhage, 1960).

(2) Jakobson, R., Fant, G., Halle, M.: "Preliminaries to speech analysis. The distinctive features and their correlates", M.I.T., Acoustics Laboratory, Techn. Rep. No. 13 (1952); 3rd printing.

G. 51-CHANNEL ANALYZER FOR SPECTRUM SAMPLING

The general specifications for the 51-channel spectrograph under construction are summarized in Fig. I-9, I-10, and I-11. A few modifications have recently been made in the plan for the particular combinations of bandwidths and frequency spacings of the filters. The group A of the filters comprises 10 filters spaced 100 c/s apart covering the frequency range 0-900 c/s. They may be used as a supplement to the main filter bank, group B, which covers a frequency range from 1000 c/s upward. The lowest of these filters, No. 11, has a constant center frequency of 1000 c/s, independently of the particular combination selected. The center frequency of the highest filter, No. 51, varies with the spacing, Δf, between successive filters. Group B of the filter bank is normally fed from a frequency transposition stage adding 1000 c/s to the incoming speech band. When used together with group A it is connected directly without this frequency translation.

The degree of overlap, defined as the ratio $B/\Delta f$ of filter bandwidth B to the spacing measure Δf is maximally 10, i.e., in combination 3, and minimally 1.25, i.e., in combinations 4 and 7. The overlap factor is 5 for combinations 2, 6, and 9, and 2.5 for combinations 1, 5, and 8. Besides these nine linear combinations there are two combinations of group B providing equal spacings on a technical mel scale.(1)

$$tm = \frac{1000\ \log(1+f/1000)}{\log 2} \qquad \text{(technical mel)}$$

The value $\Delta tm = 80$ has been selected. The total range of 3200 tm corresponds to 8200 c/s. Position 10 has constant bandwidth of 250 c/s and position 11 has constant width of tm = 300 which is 250 c/s at f = 100 c/s and 1850 c/s at f = 8200 c/s.

A reduction of the speed of the tape-recorder at the input of the filter band by a factor of 2 may be used for increasing the effective spacing and width data of the filters by the same factor. By this trick it is possible to vary the properties of the filter bank beyond the 11 combinations of Fig. I-10. The associated stretching of the time scale is a means

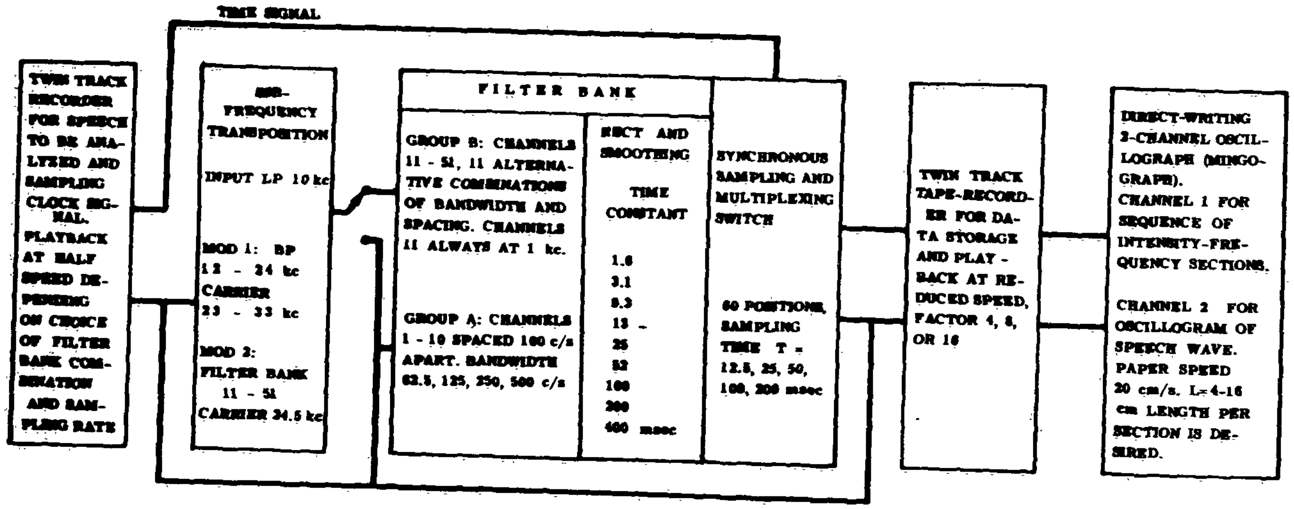
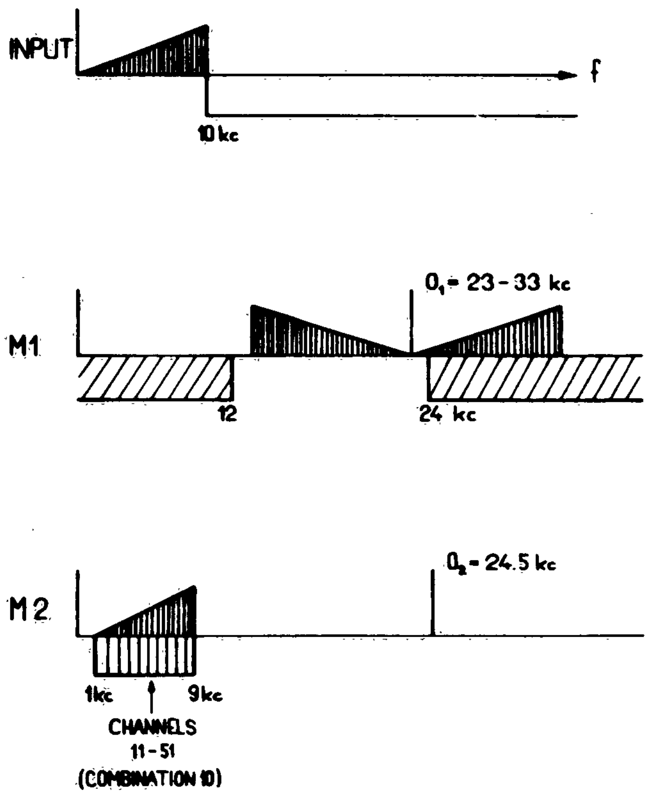

TIME SIGNAL
TWIN TRACK RECORDER FOR SPEECH TO BE ANALYZED AND SAMPLING CLOCK SIGNAL. PLAYBACK AT HALF SPEED DEPENDING ON CHOICE OF FILTER BANK COMBINATION AND SAMPLING RATE
SUB-FREQUENCY TRANSPOSITION
INPUT LP 10 kc
MOD 1: BP 12 - 24 kc CARRIER 23 - 33 kc
MOD 2: FILTER BANK 11 - 51 CARRIER 34.5 kc
FILTER BANK
GROUP B: CHANNELS 11 - 51, 11 ALTERNATIVE COMBINATIONS OF BANDWIDTH AND SPACING. CHANNELS 11 ALWAYS AT 1 kc.
GROUP A: CHANNELS 1 - 10 SPACED 100 c/s APART. BANDWIDTH 62.5, 125, 250, 500 c/s
RECT AND SMOOTHING
TIME CONSTANT
1.6
3.1
6.3
13 -
25
50
100
200
400 msec
SYNCHRONOUS SAMPLING AND MULTIPLEXING SWITCH
60 POSITIONS, SAMPLING TIME T = 12.5, 25, 50, 100, 200 msec
TWIN TRACK TAPE-RECORDER FOR DATA STORAGE AND PLAY-BACK AT REDUCED SPEED, FACTOR 4, 8, OR 16
DIRECT-WRITING 2-CHANNEL OSCILLOGRAPH (MINGOGRAPH). CHANNEL 1 FOR SEQUENCE OF INTENSITY-FREQUENCY SECTIONS.
CHANNEL 2 FOR OSCILLOGRAM OF SPEECH WAVE. PAPER SPEED 20 cm/s. L=4-16 cm LENGTH PER SECTION IS DESIRED.

Fig. 1-9 Block diagram of the 51-channel spectrum analyzer and associated equipment for the recording of successive frequency, time-synchronous spectral sections.

SSB-SYSTEM



CHANNELS 11 - 51

SPACING Δf	12.5	25	50	100	200 c/s	80 tm
RANGE W_B	0.5	1	2	4	8 kc/s	
B = BANDWIDTH						
31 c/s	1	4				
62	2					
125	3		5	7		
250			6	8		10
500				9		
300 tm						11

THE UNIT tm (TECHNICAL MEL) = $\frac{1000 \cdot \log(1 + f/1000)}{\log 2}$

WHERE f = FREQUENCY IN c/s

Fig. I-10 SSB-modulation system for filter bank group B of the 51-channel spectrum analyzer and tabulation of the 11 available filter combinations in terms of frequency spacing and bandwidth.

TYPICAL OPERATION

VARIABLE	HARMONIC ANALYSIS			FORMANT ANALYSIS			WIDE RANGE ANALYSIS		
GROUP A UTILIZED		NO		YES		NO		NO	
GROUP B CONNECTION		SSB		DIRECT		SSB		SSB	
GROUP B COMBINATION		1			9		10		11
S_1 = INPUT TAPE SPEED	1/2		1		1			1	
B = FILTER BANDWIDTH		31 kc			500 c/s		250 c/s		300 tm
B_e = EFFECTIVE BANDWIDTH	62 c/s		31 c/s		500 c/s		250 c/s		300 tm
Δf = FREQUENCY SPACING		12.5 c/s			100 c/s			80 tm	
Δf_e = EFFECTIVE FREQUENCY SPACING	25 c/s		12.5 c/s		100 c/s			80 tm	
W_B = RANGE OF GROUP B		500 c/s			4000 c/s			8000 c/s	
W_A = RANGE OF GROUP A				1000 c/s					
W_E = EFFECTIVE $(W_B + W_A)$	1000 c/s		500 c/s	5000 c/s		4000 c/s		8000 c/s	
τ = SMOOTHING TIME CONST.	50 msec		25 msec		25 msec			25 msec	
τ_e = EFF. SMOOTHING TIME CONST.		25 msec			25 msec			25 msec	
T = SWITCH PERIOD	50 msec		25 msec		25 msec			25 msec	
T_e = EFF. SAMPLING PERIOD		25 msec			25 msec			25 msec	
S_2 = OUTPUT SPEED REDUCTION	4		8		16			16	
L = PAPER LENGTH PER SAMPLE		4 cm			16 cm			16 cm	
COMMENTS					USE COMBINATION 8 OR 7 IN CASE OF LOW F_0			300 tm = 230 c/s AT 100 c/s 1900 c/s AT 8000 c/s	

Fig. I-11 Typical choice of operational characteristics for producing a sequence of frequency-intensity sections with the 51-channel spectrum analyzer.

of increasing the number of spectrum samples per second at a constant sampling rate of the switch unit. As an example it may be seen from Fig. I-10 and I-11 that the filter-spacing of 100 c/s and bandwidths of 250 c/s may be obtained either with combination 8 and the normal tape-recorder speed or with combination 5 and a speed reduction by a factor 2. Combinations 1, 2, and 4 are primarily intended for harmonic analysis whereas combinations 3, 5, 6, 7, 8, 9 are primarily intended for "broad-band" or "formant" analysis. Some of the latter will be useful in the simulation of various formant-tracking schemes.

The output from the electronic switch is the time-division multiplexed data of the rectified and smoothed outputs of the filter banks. Smoothing time constants are variable from 2.5 - 320 msec. The switch period includes 9 empty time positions besides those of the 51 channels and these may be used for the synchronous sampling of other speech parameters of interest. One application would be to check the function of a pitch frequency tracker or of a formant tracker against the short time spectrum and oscillographic display. A constant reference to be recorded on a separate channel of the oscillograph is simply the speech wave time-function. In case our direct-writing ink-jet recorder, the Mingograph, is used for the final data display it will be necessary to slow down the data flow by means of storage in a twin track tape-recorder and playback at reduced speed. This is the proposed normal practice which has the advantage of immediate accessibility of records and that large size spectrum section diagrams and time-synchronous oscillograms may be printed on light-weight, cheap paper. An alternative or rather supplementary method of spectrum portrayal would be to produce time-frequency-intensity spectrograms from a cathode ray tube.

The work is progressing on the design of a prototype for the individual channels of the analyzer. The band-pass filters contain three resonant circuits in cascade and are designed for minimum overshot characteristics. The smoothing filters are 18 dB/octave active RC-filters, also of minimum overshot type.

(1) Fant, G.: "Acoustic analysis and synthesis of speech with applications to Swedish", Ericsson Technics 15, No. 1, 3-108 (1959).

II. SPEECH SYNTHESIS AND SPEECH PERCEPTION

A. VOWEL SYNTHESIS

A set of 754 synthetic vowels have been recorded for future use in experiments on perceptional mapping of the vowel stimuli domain. Our standard procedure for series synthesis was followed.[1] This system is based on formants F1 F2 F3 F4 and a higher pole correction. Our present voice source has a zero at $p = 2\pi\cdot 800$ c/s in addition to poles at $p = 2\pi\cdot 100$, $p = 2\pi\cdot 200$, and $p = 2\pi\cdot 3000$ c/s. Formant bandwidths are held at the constant value of 70 c/s up to 1500 and constant $Q = 25$ above 1750 c/s.

J. Mártony, L. Lisker

[1] Fant, G.: "Acoustic analysis and synthesis of speech with applications to Swedish", Ericsson Technics, 15, No. 1, 3-108 (1959)

B. CONFUSIONS AMONG VOWELS FOLLOWING LOW-PASS AND HIGH-PASS FILTERING

A project directed to the study of the basis for perception of vowels has been started. Prolonged vowels from several speakers have been isolated and shaped uniformly with regard to their time duration and envelope by means of electronic gating techniques (developed by Møller). Confusion tests of the effects of low-pass and high-pass filtering on selected Swedish vowels and synthetic copies of these are planned.

E.C. Carterette

C. F-PATTERN APPROXIMATIONS OF VOICED STOPS AND FRICATIVES

Voiced stops may be approximated by the introduction of formant transitions at the onset of a vowel. In order to study the specific requirements for a series connected formant coded synthesizer we duplicated some of the Haskins Laboratories experiments on the role of the initial transitions. A number of in all 188 stimuli from the synthesizer OVE II mentioned above were recorded. Stimuli length was 300 msec with an abrupt onset of the source function and a slight fall off of the intensity during the last 50 msec of the vowel. The formant frequency transitions were given a constant slope and a duration of 25 msec in one part of the test and 50 msec in another part of the test. The steady state portion of the vowel was given a formant pattern of F_1 = 750 c/s, F_2 = 1250 c/s, F_3 = 2550 c/s, and F_4 = 3250 c/s corresponding tc the vowel [a].

The first formant started at F_1 = 150 c/s and 250 c/s and the third formant at F_3 = 2250 c/s, 2550 c/s, and 2850 c/s. The starting point for F_2 was varied in steps of 100 c/s when F_1 started at 150 c/s and in steps of 200 c/s when F_1 started at 250 c/s, within the entire F_2-range.

A preliminary listening test with two phonetically trained Swedish subjects was carried out. Only those instances when both subjects agreed that they heard a specific voiced stop or fricative with a fair degree of certainty were noted.

The phonemic responses reported were /b, d, g/ and /v, j/. All the /b, d, g/ responses were associated with the shorter transitional interval. All the v and most of the j responses were obtained in connection with the longer transitional interval. Only a few j responses occurred with the shorter transitional interval in which case the higher starting point of F_1 was a necessary requirement. However, within the group of stimuli with the longer transitional interval there was no appreciable difference in response for the higher and the lower F_1.

TABLE II-1.

F_{30} c/s	2250		2550		2850		2250		2550		2850	
F_{20} c/s \ F_{10}	150	250	150	250	150	250	150	250	150	250	150	250
500					b							
600			b	b								
700	b	-	b	-								
800	b	b	x b	b	b				v	v		
900	b	-	b	-						x -	v	
1000	b	b		b					v	v		
1100												
1200						d						
1300						-						
1400					d	d						
1500					d	-						
1600					x d	d						
1700	g				d	-					j	
1800					d	d					j	
1900	g								j		j	
2000	x g				g	j	j	j		j		x j
2100	g					-	j	-	j	-	j	-
2200						j		j		j	j	j
2300						-				-	j	-
2400						j			j	j		j
2500						-						
2600						j						
2700						-						
2800						j						

τ_1 = 25 msec τ_2 = 50 msec

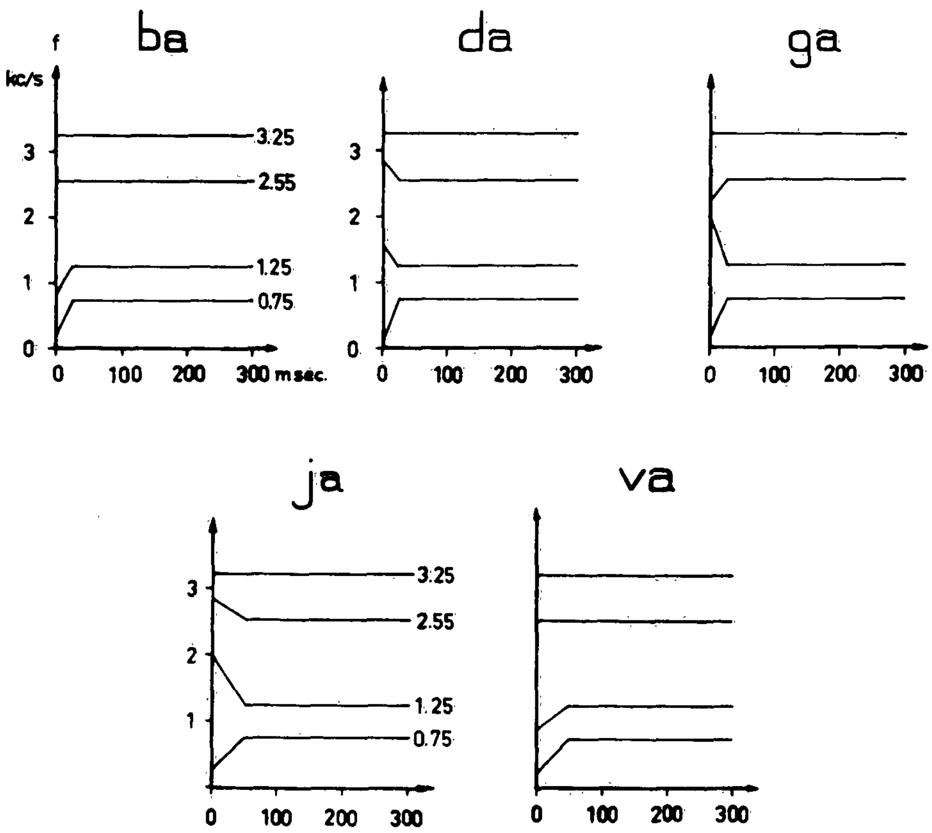


Fig. II-1 Preferred stimuli patterns for each of three voiced stops and two voiced fricatives, as judged by Swedish trained listeners.

It can be seen that an almost necessary requirement for /g/ responses is a low F_3 and a high F_2 and that the necessary requirement for /d/ responses is a high F_3 and a medium high F_2. In no instance (with the possible exception of a high F_3) was a proper choice of F_2 alone sufficient for differentiating all the stop sounds. The essential requirement for /b/ responses was a low F_2.

These results indicate that a shift in F_3 alone is sufficient for shifting the response from /g/ to /d/ and that the differentiating role of F_2 for the /g:d/ distinction is highly dependent on the particular F_3. These preliminary results are in several respects similar to those obtained at Haskins Laboratories for the pattern playback stimuli.[1] Our data show a tendency of a greater differentiating role of F3. However, both the systematic stimuli differences and language differences should be considered.

Optimal stimuli (marked with x in Table II-1) patterns within our simplified inventory are shown in Fig. II-1.

L. Lisker, J. Mártony, B. Lindblom, S. Öhman

[1] Harris, K.S., Hoffman, H., Liberman, A.M., Delattre, P.C., Cooper, F.S.: "Effect of third-formant transitions on the preception of the voiced stop consonants", J.Acoust.Soc.Am. 30, 122-126 (1958).

III. SPEECH PRODUCTION

A. X-RAY DEMONSTRATION FILM

An X-ray demonstration film illustrating coarticulation effects has been completed. The major part of the work was undertaken in cooperation with the Wenner-Gren Research Laboratory, Stockholm.

H.M. Truby

B. X-RAY TECHNIQUES APPLIED TO THE STUDY OF NASALIZATION

The Speech Transmission Laboratory has participated in a project on cleft palate speech directed by B. Nylén (plastic surgery specialist) and L. Björk (X-ray specialist). Our contribution has been the introduction of Visible Speech spectrographic techniques for objective studies of the effects of various types of nasalization and methods of synchronizing X-ray moving film with a timing signal which is recorded on one track of the tape-recorder. One part of the investigation is concerned with the movements of the velum in normal speech in relation to the visible effects on spectrograms. The study is continuing.

L. Björk, A. Møller, B. Nylén

STL-QPSR-2/1960 (October-December)

Speech Transmission Laboratory

QUARTERLY PROGRESS AND STATUS REPORT

January 15, 1961

SPEECH TRANSMISSION LABORATORY
Div. of Telegraphy-Telephony
Royal Institute of Technology
Stockholm/Sweden

STL-QPSR-2/1960 (October-December)

Speech Transmission Laboratory

QUARTERLY PROGRESS AND STATUS REPORT

January 15, 1961

SPEECH TRANSMISSION LABORATORY
Div. of Telegraphy-Telephony
Royal Institute of Technology
Stockholm/Sweden

ACKNOWLEDGMENTS

The research reported in this document was made possible in part by supports from the following sources:

		Chapter
State Council of Technical Research (Statens Tekniska Forskningsråd)	(1)	I,II,III, IV
State Council of Medical Research (Statens Medicinska Forskningsråd)	(4)	I D
Research Institute of the National Swedish Defence (Försvarets Forskningsanstalt)	(3)	I,II
Knut och Alice Wallenbergs Stiftelse	(4)	I,III,IV
US Air Force under Contract AF 61(052)-342	(3)	I,II,III
US Army under Contract DA-91-591-EUC-1496 OI-4676-60	(2)	I

(1) Speech communication research on a broad basis
(2) Speech research, particularly speech analysis techniques
(3) Speech research, particularly speech analysis and synthesis
(4) Speech and hearing research of medical interest

TABLE OF CONTENTS

SPEECH TRANSMISSION LABORATORY x)

PERSONNEL

ADMINISTRATION AND PERMANENT STAFF

Speech Communication Research:

Tekn.dr., Doc. G. Fant	Director of the Speech Transmission Laboratory and the Speech Communication Research Group
Fru Marianne Richter	Secretary
Fru Si Felicetti	Secretary
Civ.ing. U. Rengman	Assistant manager and research associate
Ing. B. Wejnebring	Head of laboratory workshop
Civ.ing. A. Risberg	Assistant director of Speech Communication Research Group

Technical Audiology:

Ing. B. Johansson	Head of the Technical Audiology Group
Fru Ester Lindberg	Secretary

Hearing Research:

Dipl.ing. G. Rösler	(Karolinska Institutet, Fysiologisk Akustik)
Herr E. Voolahe	Audiometry technician

RESEARCH ASSOCIATES

Med.kand., Civ.ing. C. Cederlund	fellowship from State Council of Technical Research
Civ.ing. J. Liljencrants	
Fil.mag. B. Lindblom	
Dipl.ing. (E.T.H.) J. Mártony	
Ing. A. Møller	
Fil.dr. H.M. Truby	
Fil.kand. S. Öhman	

x) The Speech Transmission Laboratory is a research department within the Division of Telegraphy-Telephony (Head Professor Torbern Laurent)

GUEST RESEARCHERS AND TEMPORARY ASSOCIATES

Dr. E.C. Carterette	(Assistant professor of psychology, University of California, Los Angeles, USA) National Science Foundation Postdoctoral Fellow
Mr. J.N. Holmes	(Joint Speech Research Unit, Post Office, Eastcote, England)
Siv.ing. A. Krokstad	(Norges tekniske högskole, Trondheim, Norge) fellowship from Scandinavian Council for Applied Research
Siv.ing. M. Kringlebotn	(Norges tekniske högskole, Trondheim, Norge) fellowship from Scandinavian Council for Applied Research

TECHNICAL ASSISTANTS

Herr L. Andersson x)
Herr S.-E. Appelgren x)
Herr S. Berg
Herr B. Lindström
Fru Ingrid Møller

THESIS STUDENT

Teknolog G. Garpendahl

PERSONNEL FROM INSTITUTIONS OUTSIDE THE R.I.T. CONTRIBUTING TO THE WORK DURING THE PERIOD OF REPORT

Med.lic. L. Björk	Akademiska Sjukhuset, Uppsala
Med.lic. B. Nylén	Akademiska Sjukhuset, Uppsala
Med.dr., Doc. E. Wedenberg	Karolinska Institutet, Stockholm

x) Technical Audiology Group

STATUS OF PUBLICATIONS AND REPORTS [x)]

		status:
Fant, G.:	"Acoustic Theory of Speech Production" Mouton & Co., 's-Gravenhage 1960, 323 pp.	in print
Fant, G.:	"The Acoustics of Speech" invited paper presented at the 3rd International Congress on Acoustics, Stuttgart September 1959, to be publ. by Elsevier Publ.Co.,Amsterdam, in the Proceedings from this congress	
Fant, G.:	"Descriptive Analysis of the Acoustic Aspects of Speech" invited paper presented at the Wenner-Gren Foundation for Anthropological Research Symposium on Comparative Aspects of Human Communication at Burg Wartenstein/Austria September 1960, to be publ. in LOGOS the Bulletin of the National Hospital for Speech Research	
Fant, G.,Stevens, K.N.:	"Systems for Speech Compression" Fortschritte der Hochfrequenztechnik Vol. 5 (1960) 229-262, Akademische Verlagsgesellschaft M.B.H., Frankfurt am Main	in print
Johansson, B.:	"A New Coding Amplifier System for the Severely Hard of Hearing" paper presented at the 3rd International Congress on Acoustics, Stuttgart, Sept. 1959, to be publ. by Elsevier Publ.Co.,Amsterdam, in the Proceedings from this congress	
Møller, A.:	"Improved Technique for Detailed Measurements of the Middle Ear Impedance" J.Acoust.Soc.Am. 32, 250-257 (1960)	
Møller, A.:	"Network Model of the Middle Ear" to be publ. in J.Acoust.Soc.Am. 33, Febr. 1961	
Møller, A.:	"The Acoustic Impedance of the Human Ear" paper presented at the 3rd International Congress on Acoustics, Stuttgart, Sept. 1959, to be publ. by Elsevier Publ.Co.,Amsterdam, in the Proceedings from this congress	
Wedenberg, E.:	"Auditory Training of the Severely Hard of Hearing Using Coding Amplifier", paper presented at the 3rd International Congress on Acoustics, Stuttgart, Sept. 1959, in the Proceedings from this congress	

[x)] All publications and reports which are under preparation or which have recently been published are included in this list.

INTRODUCTION

This report series is intended for quarterly summaries of recent progress and status of research at the Speech Transmission Laboratory of the Royal Institute of Technology, Stockholm. The present issue, STL-QPSR-2/1960, is the second report of the series intended for general distribution. Active projects not reported on here or mentioned briefly only were dealt with in greater detail in the first issue, STL-QPSR-1/1960. Any particular results reviewed here are to be considered as preliminary only and will normally reappear in coming scientific papers.

Stockholm, January 15, 1961
Gunnar Fant

I. SPEECH ANALYSIS

A. SPECTRUM SAMPLING INSTRUMENTATION

1. The 51-channel analyzer

Detailed design plans for the 51-channel analyzer were given in the previous progress report. The design of the individual channels of the filterbank is almost completed. At present we are investigating temperature stability and other drift sources and a scheme for tuning the coils and condensers of the filters. The constructional phase will be started in about 3-4 months' time.

J. Liljencrants, U. Rengman

2. RASSLAN - a 6-channel closed loop sectioning device

A provisional instrumentation for spectrum sampling has been constructed, to a large extent utilizing earlier existing units. It has a comparatively low data handling capacity but can do a certain amount of routine work till the 51-channel spectrometer is completed.

The general design and function of the analyzer are illustrated in Fig. I-1.

The speech sample to be analyzed is stored on a tape loop with a revolution time of the order of 2 sec. On the loop is also a triggering signal, at present a 10-kc/s tone burst with a duration of 10 msec. This signal actuates a manually adjustable delay unit, the output of which governs the sampling point. The speech signal is processed in a heterodyne system analogous to that of the proposed 51-channel analyzer. The band-pass analysis is carried out in a bank of six filters with center frequencies from 1 kc/s and upwards. The output of the filters are individually rectified and smoothed. At the arrival of the delayed triggering pulse these signals are simultaneously sampled and stored in memory capacitors. The voltages of these are successively transferred to the recording system with the aid of a

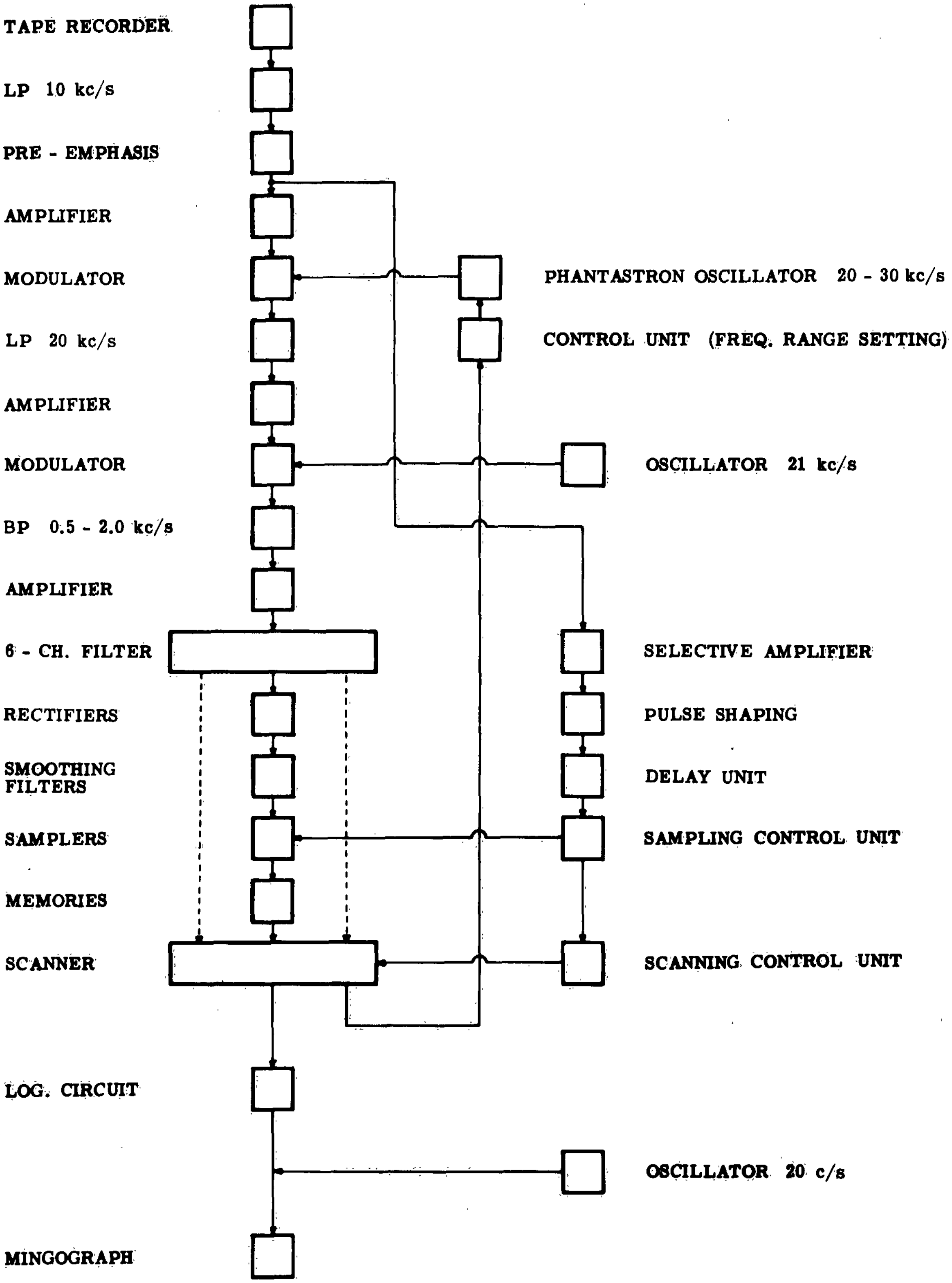


Fig. I-1 Block diagram of the six-channel closed loop spectral sectioning instrumentation.

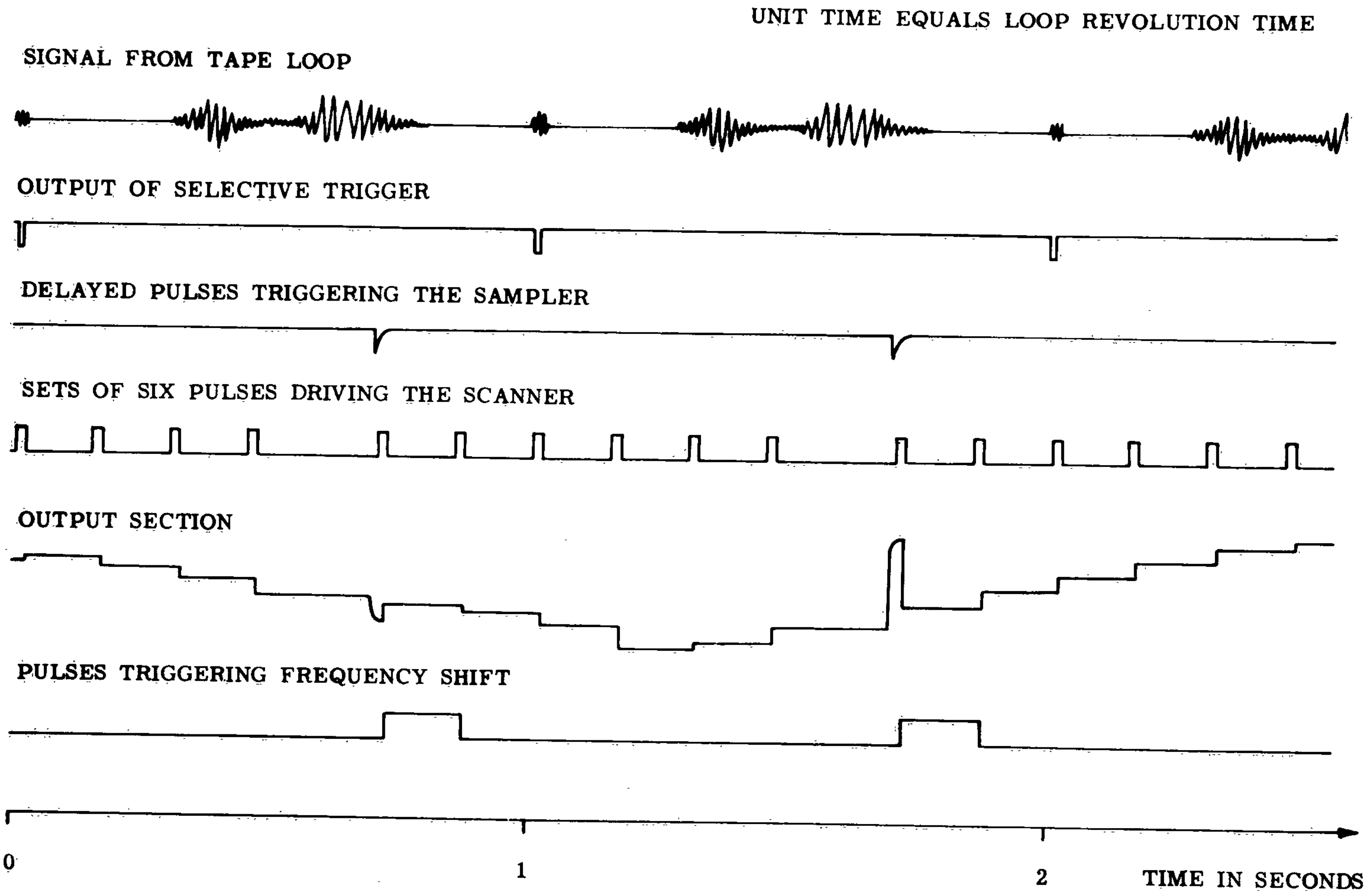

UNIT TIME EQUALS LOOP REVOLUTION TIME
SIGNAL FROM TAPE LOOP
OUTPUT OF SELECTIVE TRIGGER
DELAYED PULSES TRIGGERING THE SAMPLER
SETS OF SIX PULSES DRIVING THE SCANNER
OUTPUT SECTION
PULSES TRIGGERING FREQUENCY SHIFT
0
1
2
TIME IN SECONDS

Fig. I-2 Sampling diagram for the six-channel spectrograph.

rotating telephone switch with a revolution time equal to that of the tape loop. When the synchronous sampling is carried out a 30-position telephone switch is put into action. This switch moves on one step selecting a carrier control voltage appropriate for the next cycle of analysis. For each revolution the frequency range of analysis is thus shifted by a constant amount. A complete spectrum section is obtained after 30 revolutions of the tape loop, a procedure that will take about 60 sec. The output of the scanner is converted to a logarithmic measure with the aid of a copper monoxide diode with forward current. The recording is finally carried out on a Mingograph recorder operating at a low paper speed (5 to 20 mm/s). In order to give a more clear picture the recorded signal is superimposed on a constant, low-frequency sine wave. The section consists of 30 adjacent sets of 6 columns representing the spectral energy in 180 bands with equal spacing and width.

The six analyzing filters have adjustable center frequencies and bandwidths corresponding to spacing of 15, 25, 50, and 100 c/s and bandwidths of 31, 62, 125, and 250 c/s, respectively. With the four mentioned spacings a total frequency range of 2.7, 4.5, 9.0, and 18.0 kc/s, respectively, will be covered. The heterodyne system, however, can manage frequencies below 10 kc/s only. Higher frequency ranges may be reached by means of reducing the playback speed of the sounds on the loop The absolute position of the frequency range of the analysis may be chosen arbitrarily.

The smoothing filters have integration times variable in octave steps from 10 to 160 msec. All filters have the STL standard minimum overshoot characteristics.

As illustrated in Fig. I-2 there will occur an extra spike between every sixth column in the output section owing to the recharging of the memory capacitors. These spikes are helpful for determining the frequency scale. The dynamic range is somewhat wider than 40 dB.

In Fig I-3 and I-4 are shown some typical analysis records from RASSLAN. It should be noted that these sections are taken from signals with high-frequency pre-emphasis (pole at $-2\pi \cdot 5000$ r/s, zero

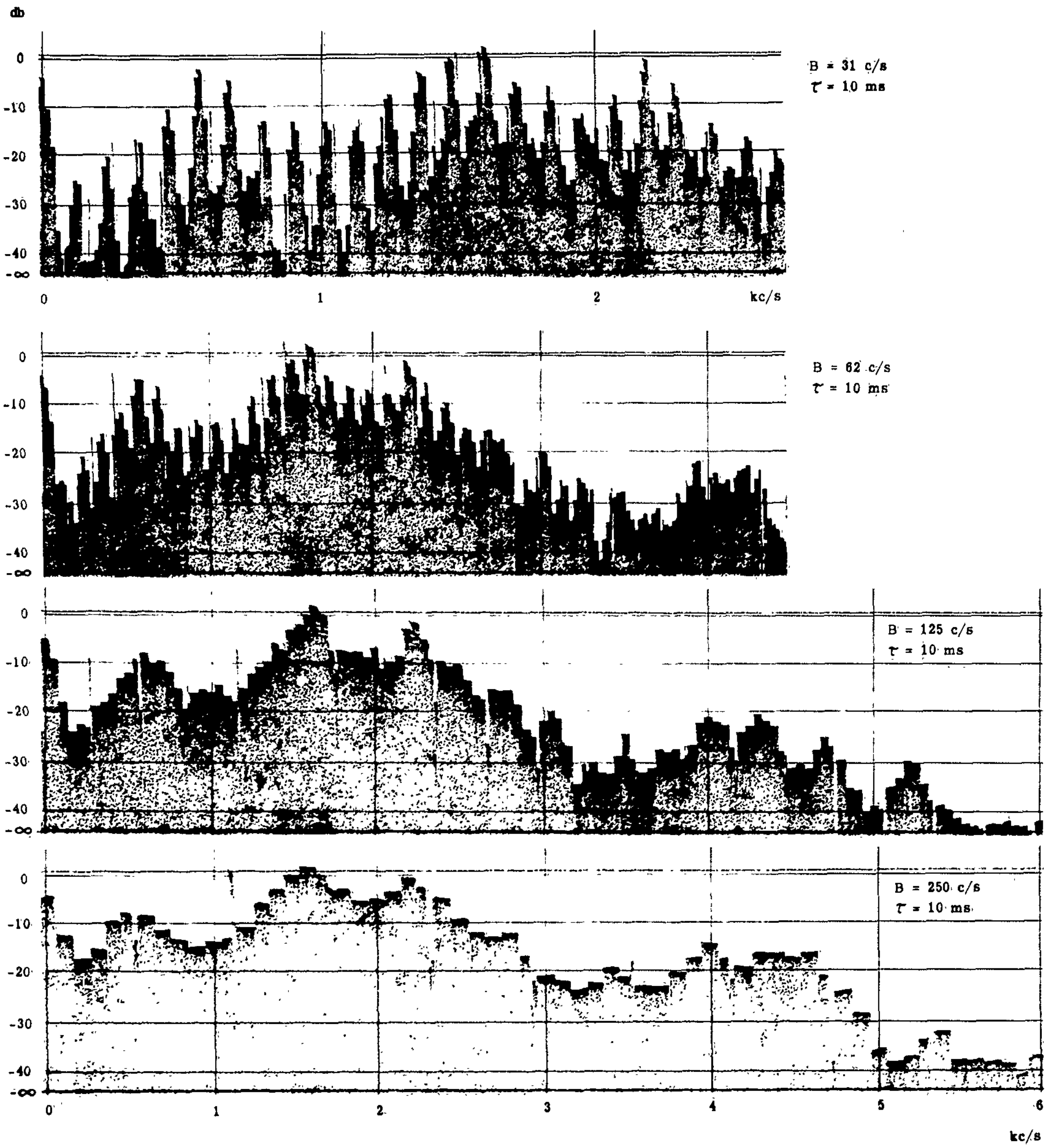


Fig. 1—3 Examples of spectral sections of a vowel [æ] produced with the six-channel spectrograph.

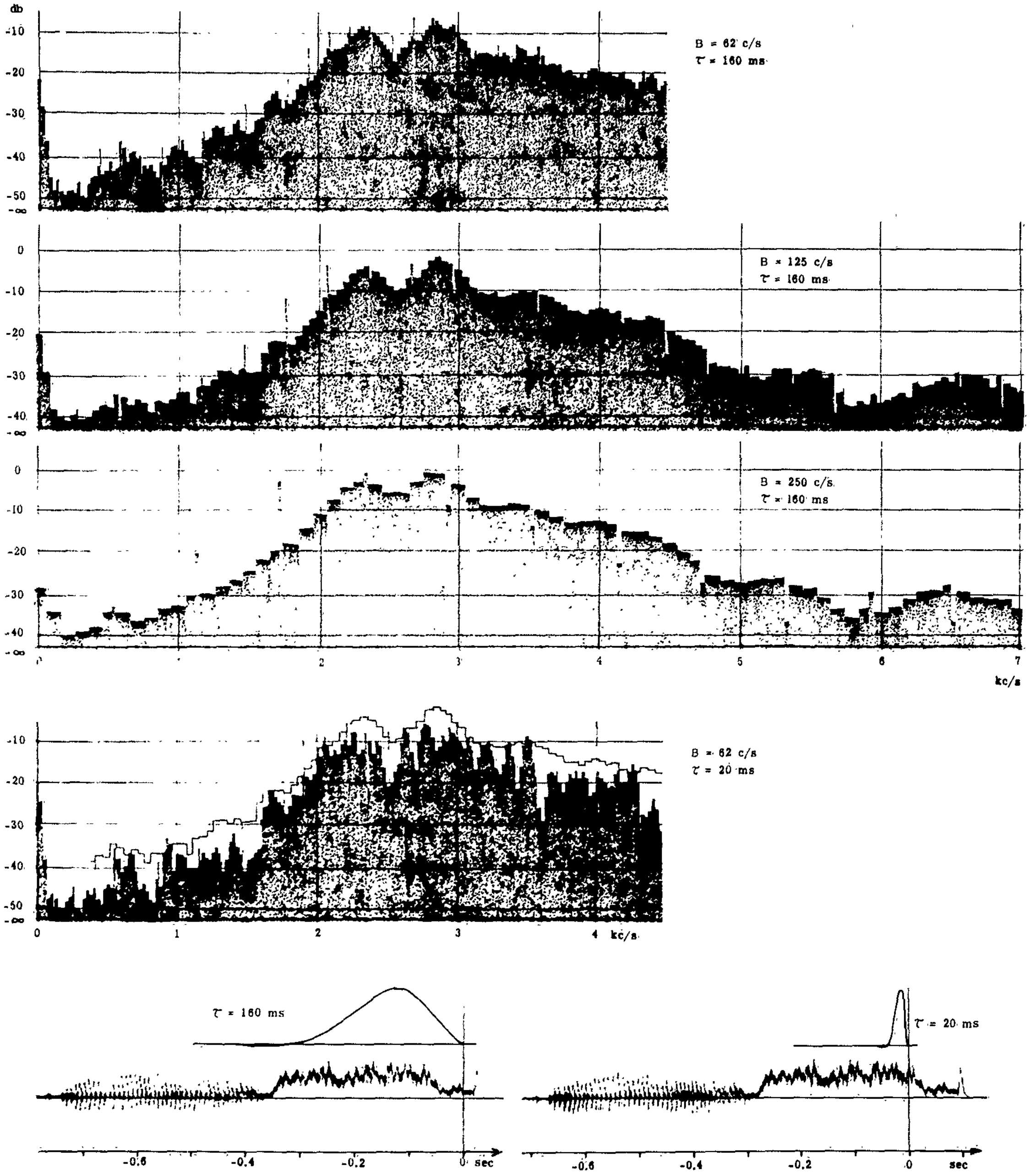


Fig. 1—4 Examples of spectral sections of a fricative [ʃ] produced with the six-channel analyzer.

at $-2\pi\cdot 200$ r/s).

The narrow-band analysis of the vowel (B=31 c/s and 62 c/s) displays an accurate harmonic structure the envelope of which contains more details than the wide-band spectrum (B=250 c/s). The analysis of the fricative, Fig. I-4, illustrates the influence of the quotient between analysis and smoothing filter bandwidths upon the random deviations from the mean envelope, see section A-3. Below the spectral section there is included a duplex oscillogram of the speech wave time function and a memory function for the smoothing filter, the time base of which is located at the sampling point.

J. Liljencrants

3. Analysis of random signals

The spectral analysis of random noise signals involves a band limitation of a band-pass filter of width B followed by a rectification and smoothing through a low-pass filter having the integration time τ.* The DC voltage thus obtained (V) has a random ripple superimposed on it, the RMS value of which may be called σ_V.

It is known [(1)] that

$$\frac{\sigma_V}{V} = k\cdot \frac{1}{\sqrt{B\tau}} \qquad \text{(Eq. I-1)}$$

See for instance ref. (1) and the previous progress report. If the signal is plotted on a logarithmic scale and provided σ_V:V is reasonably small this formula can be rewritten:

$$\sigma_y = k\cdot 20\cdot \log_{10} e\cdot \frac{1}{\sqrt{B\tau}} \qquad \text{(Eq. I-2)}$$

which is the RMS value of random striations in dB. The conversion

* For definition of τ see ref. (2). In terms of equivalent bandwidth $\tau = 1/2F$, where F is close to the 6 dB cutoff frequency of the LP-filter.

from a linear scale to a dB-scale will cause a shift in the mean level, but owing to the single curvature of the logarithmic function the approximation above is still fairly good. A variation $\Delta V{:}V = \pm 50\ \%$ will, according to the formula, give $\Delta Y = \pm 4.4$ dB while the correct values are $\Delta Y = {}^{+3.5}_{-6.0}$ dB. Consequently the approximative value is some 10 % too low.

σ_y is generally smaller than mentioned in this example. An experiment using our standard third order minimum overshoot filters for both band-pass filtering and smoothing has given an empirical value of the constant:

$$\sigma_y = \frac{3.45}{\sqrt{B\tau}}\ \text{dB} \qquad \text{(Eq. I-3)}$$

where B is the analyzing filter bandwidth in c/s and τ is the integration time of the smoothing filter in seconds.

An illustrative example:

In Fig. I-4 one section is taken from a [ʃ]-sound with B=62 c/s and τ=20 msec (F=20 c/s). According to Eq. I-3 these values correspond to $\sigma_y = 3.1$ dB. In the same diagram is plotted the result from an analysis with B=125 c/s and τ=160 msec. The latter section has a higher level owing to the larger value of B and the different time span of the sample. A calculation of the standard deviation apart from this overall shift gives the value $\sigma_y = 3.8$ dB. This value is only slightly larger than could be expected from random variations in the two samples indicating that the spectral shape of the sound during the shorter interval is essentially the same as the "mean shape" of the sound.

J.N. Holmes, J. Liljencrants

(1) Morrow, C.T.: "Averaging time and data-reduction time for random vibration spectra", Part I in J.Acoust.Soc.Am. 30, 456-461 (1958); and Part II in J.Acoust.Soc.Am. 30, 572-578 (1958).

(2) Fant, G.: "Acoustic analysis and synthesis of speech with applications to Swedish", Ericsson Technics, 15, No. 1, 3-108 (1959).

B. FORMANT FREQUENCY MEASUREMENTS

1. Spectrographic measurements

5 subjects were asked to determine the positions of F_1 F_2 and F_3 of 6 synthetic vowels [a:, e:, i:, ʉ:, oe:, ae:] on wide-band spectrogram and on wide-band section. Each vowel has 6 pitches. This gives a total of 540 measurements for each method. Mean values of errors in each F_0-class are shown in Fig. I-5, where the lower diagram pertains to the mean of absolute values of errors. An error is simply defined as the deviation of a measured value from the known pole frequency.

Analysis of high-pitched voices with 600 c/s filter, which implies a reduction of the speed of the input signal by a factor of 2 and the use of a magnified frequency scale, produces a spectrographic display which makes accurate measuring very difficult (cf. table I-1). Moreover, the effect of the 6 dB/octave bass attenuation is particularly noticeable in this method. In a systematic investigation subjects consistently located low formants much too high, e.g., F_1 of [i:] at about 400 c/s instead of 240 c/s pole.

TABLE I-1

	section		spectrogram		48-channel
analysis filter	45	300	300	600	350 c/s
average error	40	55	50	90	50 c/s
spread	-	60	70	150	- c/s
maximal error	90	170	150	250	200 c/s

factors influencing measurements:

(1) uncertainty in calibration, position of zero line

(2) pre-emphasis: +6 dB/octave (of importance in the case of 600 c/s filter)

(3) position of partial within the formant (see Fig. I-6)

(4) relation between filter width and fundamental pitch.

The effect of a continuous rise in F_0 on the appearance of human and synthetic vowels may be studied in Fig. I-6.

In order to perform an auditive evaluation of the above-mentioned wide-band (300 c/s) spectrogram measurements a preliminary listening test has been carried out. Subjects were asked to state

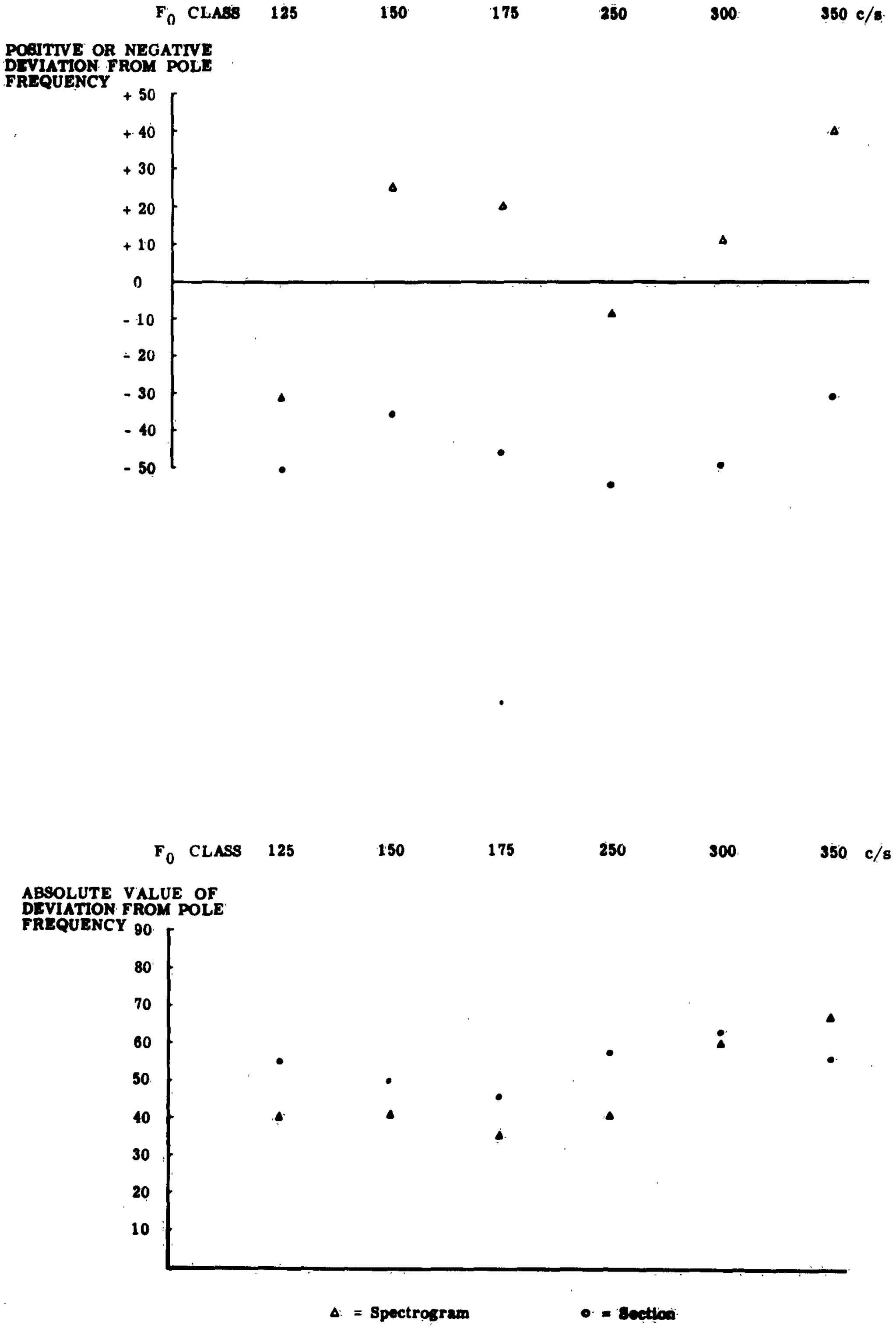


Fig. I-5 Mean values of errors in spectrographic measurements, with signs retained (above) and means of absolute values (below). Triangles pertain to spectrograms and circles to wide-band sections.

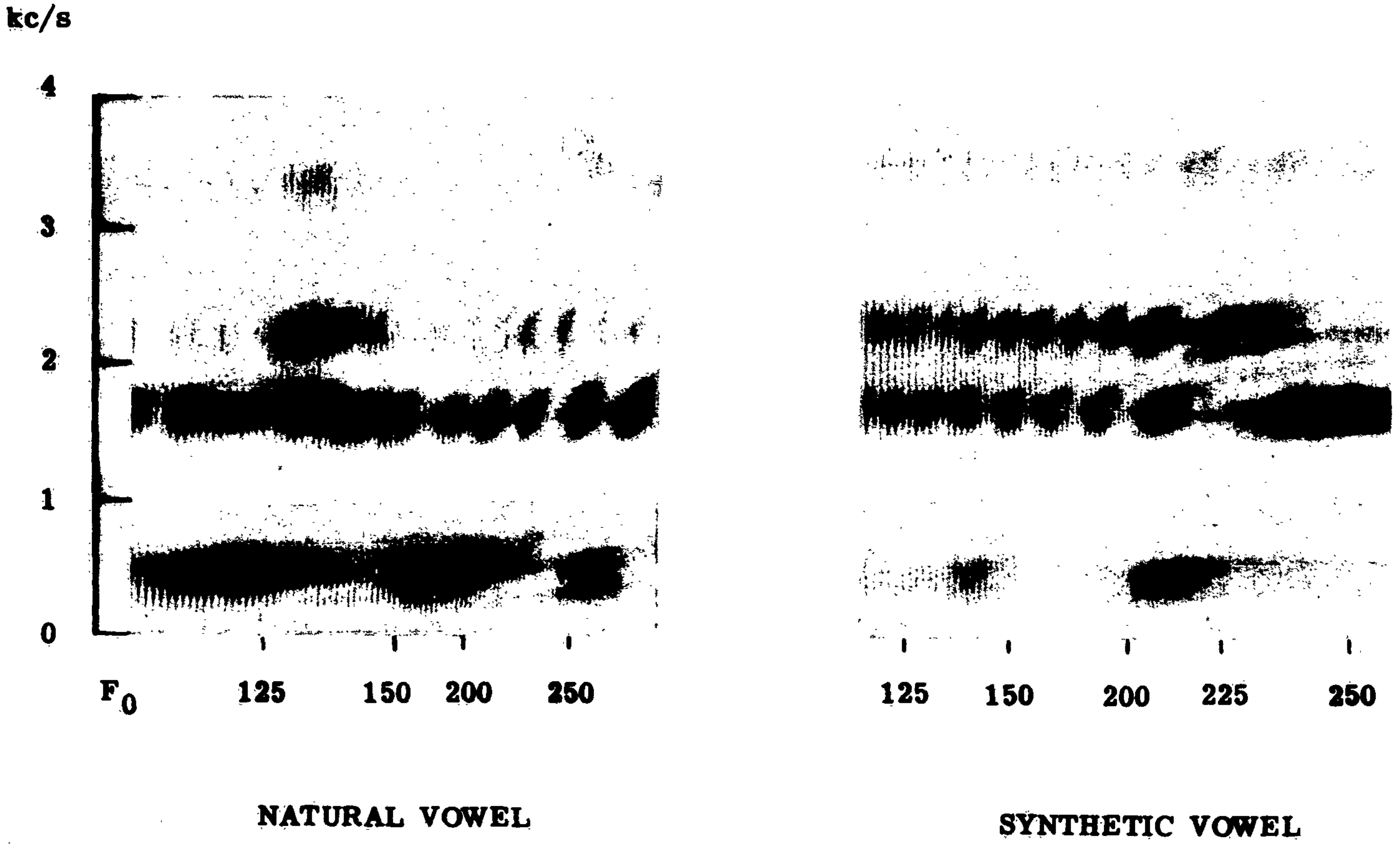


Fig. I-6 The effect of a gliding pitch change on broad-band spectrogram of a vowel [æ] . The human speech and the synthetic speech sample show similar jumps in the locations of the formant bands.

whether pairwise stimuli consisted of sounds that were the same or different with respect to phonetic quality. A pair contained two synthesized vowels; one of the sounds used for the above-mentioned spectrogram study plus its measured counterpart specified by the spectrogram data. Subjects were unanimous in reporting that one or two out of 36 sounds differed considerably whereas the rest had remained practically unchanged with respect to phonetic quality.

B. Lindblom

2. Formant-tracking

A few pilot experiments on the use of phase detection methods in formant frequency trackers have been undertaken. The methods we have tried were:

(1) Measuring the phase difference between the input and the output of an anti-resonance circuit of variable center frequency.

(2) Measuring the phase difference between the outputs of two anti-resonance circuits spaced 250 c/s apart.

(3) Measuring the phase difference between the outputs of two resonance circuits spaced 100 c/s apart.

Of these methods only the last one gave results which were judged to be of any practical use. However, a large pitch variation alone could give a phase shift of a formant frequency. The results were better for F_2 and F_3 than for F_1. No further work on phase methods will be undertaken until sufficient experience has been gained from spectral maximum selectors.

A. Krokstad

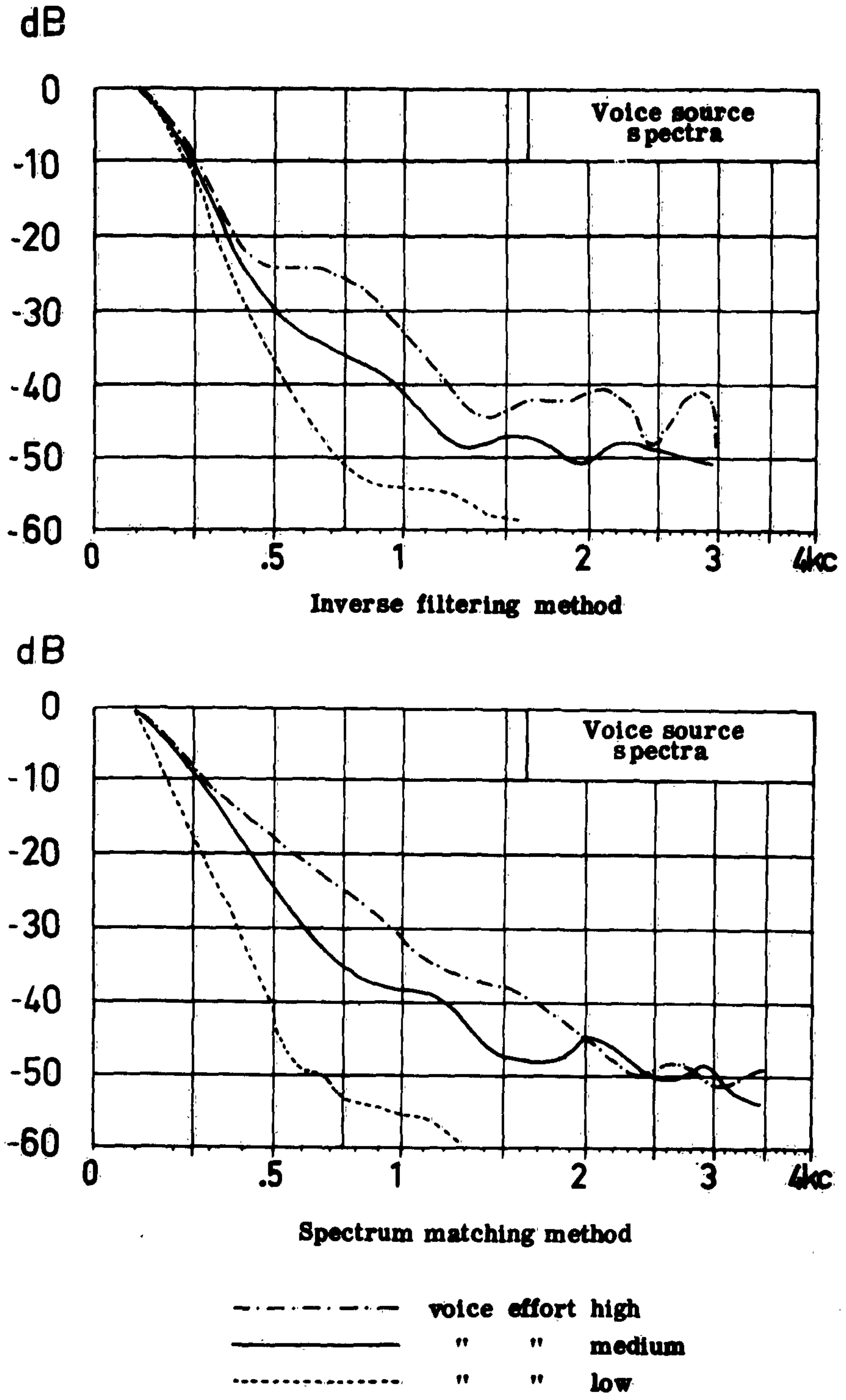
C. POLE-ZERO MATCHING TECHNIQUES

The work on pole-zero matching of fricatives and vowels exemplified in the previous progress report has been continued. Part of the work is intended to serve as the basis for the synthesis of connected English speech. Another part of the work is devoted to the pole-zero specification of vowel spectra as reported in section D below.

Several techniques have been tried. Besides the procedure utilized in the previous work comprising a graphical summation of the contributions from individual complex poles and zeros, i.e., elementary resonance and anti-resonance curves, we have tried a similar summation of the contributions to the spectrum envelope of each pole or zero, positive and negative parts of a complex pole treated separately. This latter method has the advantage of simplicity in the graphical work but is generally not as fast as the method with elementary resonance curves. Among other methods looked into recently is the use of an electrolytic tank with accessories for frequency response tracing.

The most convenient method from an instrumental point of view is the use of analog equipment. The inverse or "anti-resonance" filtering method, which we have relied on to a large extent for voice source studies, represents a time-domain technique. It gives good accuracy for measuring F_1 and F_2 of vowels and could be supplemented by other methods of a complete pole-zero matching of the vowel sample. The use of synthesis instrumentation for producing synthetic spectra that are compared with those of human origin should be quite practical providing the circuitry may be handled and calibrated with sufficient accuracy and with sufficient ease of operation. The present circuitry in OVE II for generating fricative sounds does not altogether fulfill these requirements and contains only two poles and one zero. Synthesis circuitry for pole-zero matching employing three additional pole-zero pairs is under construction, see section II-B of this report. One of the advantages of the analog circuitry approach for pole-zero matching is that the perceptual significance of various approximations may be tested by a synthesis experiment employing the same circuitry.

C. Cederlund, J.N. Holmes, M. Kringlebotn, J. Mártony



Inverse filtering method

Spectrum matching method

Fig. I-7 Voice source spectra at three different voice efforts in producing the vowel [æ]. The overall VU-level of the vowels differed in steps of 10 dB. The top diagram is the result of inverse filtering analysis and the bottom diagram is the result of spectrum matching (analysis by synthesis).

D. VOICE SOURCE STUDIES

The continued studies of the properties of vocal sound sources have concentrated on extraction of the spectrum envelope of the voice source. Two methods have been tried. One is the frequency analysis of the output of the inverse-filtering circuitry after a maximally accurate compensation of the time-domain formant ripple. The other involves the graphical calculation of the difference between the spectrum (broad-band section) of the sound and the spectrum of a synthetic copy given the same formant frequencies and a standardized voice source and a higher pole correction.

The primary object of the analysis is the complete source spectrum. A pole-zero matching of typical results is also attempted. Results will be given in later reports.

Examples of voice source spectra obtained by the two methods are shown in Fig. I-7. Samples pertaining to a speaker's high, medium, and low voice efforts were analyzed by the two methods. The typical effects of the relative increase of the spectrum level at higher frequencies as a result of increasing voice effort is seen. The two methods provide approximately the same results but there is a spread of a couple of dB to consider.

The accuracy of analysis decreases in the frequency range above F2 and at frequency intervals of a low spectrum level. The upper frequency limit for the significance of the data is of the order of 3000 c/s. The particular higher pole correction chosen for the synthesis affects the validity of the upper frequency range. For synthesis work, however, this is of less concern.

C. Cederlund, J. Mártony

E. VOICE FUNDAMENTAL FREQUENCY TRACKING

Work has continued on the pitch-tracking scheme described in the previous report. Three complete pitch-measuring channels have been constructed. These are separated at the input end by means of band-pass filters covering the FO-range 60-300 c/s and spaced 0.8 octave apart. The outputs of the three systems are connected to a minimum detector for selecting the lowest of the three independently measured pitch values. Experiments with a frequency separation of the three channels by means of low-pass filters instead of band-pass filters gave less satisfactory results. There is also evidence that four band-pass filters of a more narrow bandwidth, e.g., 0.6 octave, should be utilized.

The frequency-measuring unit in each channel comprises circuitry for producing zero-crossing pulses of constant amplitude and shape. The low-frequency channel incorporates a full-wave rectifier for increasing the number of zero-crossing pulses from 2 to 4 per pitch period. The final integration of the pitch frequency in each channel is performed by means of third order minimum overshoot low-pass filters. The threshold amplitude gating signals are controlled by the same type of filters. Examples of the pitch analog signals in each cf the three channels and in the output of the minimum selector are shown in Fig. I-8. The function of the selector is satisfactory but there occasionally enters a frequency ripple up to 5 c/s in the output of the second channel. This and other effects will be looked into.

A voice fundamental frequency emphasizing circuit intended as an input stage to the pitch tracker is under construction. It is based on the inverse-filtering method for reducing the relative amplitude of the first formant oscillation. The system comprises the combination of an F_1-formant tracker based on maximum selection and a gating system for enabling essentially one of a number of 8 transmission paths in parallel containing anti-resonance filters of the same center frequencies as the corresponding band-pass filters of the F_1-selector. This instrumentation has given promising results but is not quite completed.

A. Risberg

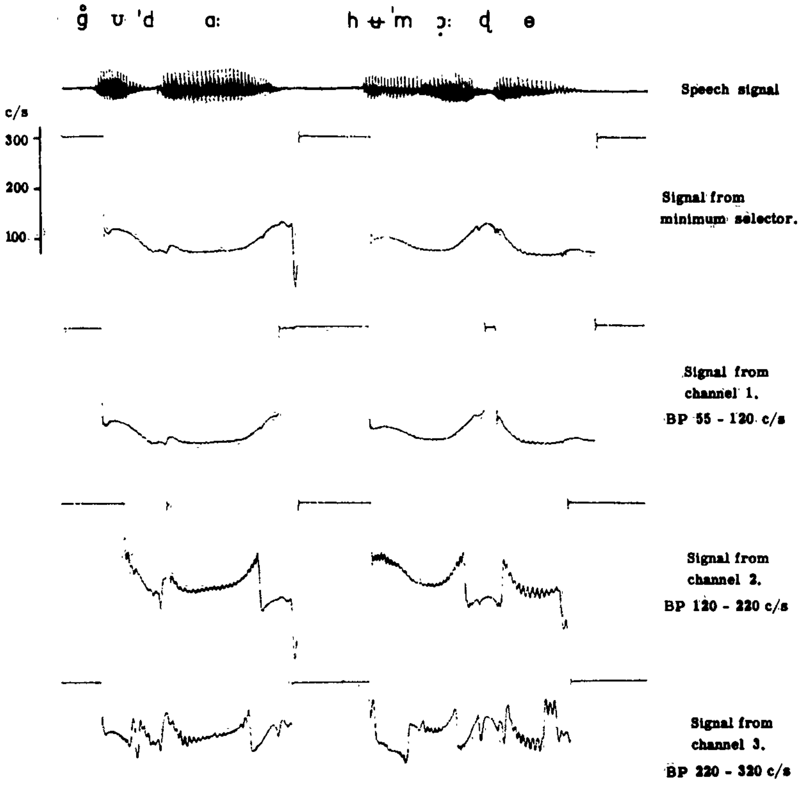


Fig. 1—8 Oscillographic test of the minimum frequency selection in the three-channel pitch tracker.

F. AUTOMATIC IDENTIFICATION OF SOUND FEATURES

A few pilot studies devoted to the identification of nasal consonants in Swedish speech material have been started. No conclusive results may be reported yet. Measurements with the aid of anti-resonance circuitry did not support the view of a rather large first formant bandwidth of nasal sounds compared with other sounds such as [l] [j] [v] and vowels of a very low F_1.

C. Cederlund, B. Lindblom, J. Mårtony, A. Møller, S. Öhman

G. STRUCTURAL CLASSIFICATION OF SWEDISH PHONEMES

The following tabulation of Swedish phonemes in a distinctive feature code is in all essentials based on the system of Jakobson-Fant-Halle (1) but with the modifications owing to the general advance of the theory and the specific views held by the author. The particular solution for the vowel system is the same as that proposed in recent publications (2), but differs from that of an earlier study (3). The consonant system has not been published before.

The distinctive feature scheme primarily serves the purpose of linguistic theory but includes acoustic descriptions which theoretically could be regarded as an instruction for machine recognition of spoken text. In practice the acoustic definition of the features often involves stipulations concerning the differences between alternative phonemes. These abstractions may not be translated to identification rules without taking into account the specific range of qualities utilized by the particular speaker in a specific context.

Most of the distinctive features or rather "phonemic distinctions" are identical with elementary phonetic categories which are well established in linguistic theory. Any scheme for machine recognition of spoken items will to some extent rely on a classification in terms of these categories. On the other hand, it is clear that an optimal recognition process will differ from the traditional distinctive feature system in terms of the particular choice and definitions of features and the sequence of operations. Vowels could thus be identified directly from properly normalized formant frequencies. Independently of the purpose of the identification scheme, whether this be linguistic theory or machine recognition of speech, it should be recognized that distinctive features or the several cues which may underlie a distinction are not always static constituents of a single sound segment[*] but quite often involve several adjacent segments and dynamic relations within a sequence of segments. These are well established facts.[**]

The Swedish vowel system comprises nine long vowel phonemes /o$_1$/ /å$_1$/ /a$_1$/ /y$_1$/ /u$_1$/ /ö$_1$/ /i$_1$/ /e$_1$/ /ä$_1$/ and nine short vowel phonemes /o$_2$/ /å$_2$/ /a$_2$/ /y$_2$/ /u$_2$/ /ö$_2$/ /i$_2$/ /e$_2$/ /ä$_2$/. These phonemic notations of the STA-alphabet conform with common Swedish orthography. Phonetic values of the basic allophones are indicated in Fig. I-9 and I-10.

The relation of /ä$_1$/ to /e$_1$/ or of /e$_1$/ to /i$_1$/ specified by the compactness feature is that of an open versus a close vowel, referring to the mouth cavity. The relation of /a$_1$/ to /å$_1$/ or /å$_1$/ to /o$_1$/ may either be identified by the compactness feature as in Fig. I-9 or with the relation of unrounded to rounded (flat) vowels, Fig. I-10. There are of course hybrid alternatives, such as the labeling of /o$_1$/ to /å$_1$/ as flat versus /a$_1$/ and opposing /å$_1$/ to /o$_1$/ in terms of compactness. This solution avoids the use of ± terms within the back vowels. The same relations hold for short vowels.

[*] The difference between the prosodic and inherent features are thus to some extent eliminated since both involve temporal relations. The prosodic features, however, generally operate over speech-wave units of a greater length than the inherent features.

[**] An attempt to construct a scheme of segment classification according to an inventory of narrow phonetic categories instead of phonemic distinctions has been undertaken in a recent article.[4]

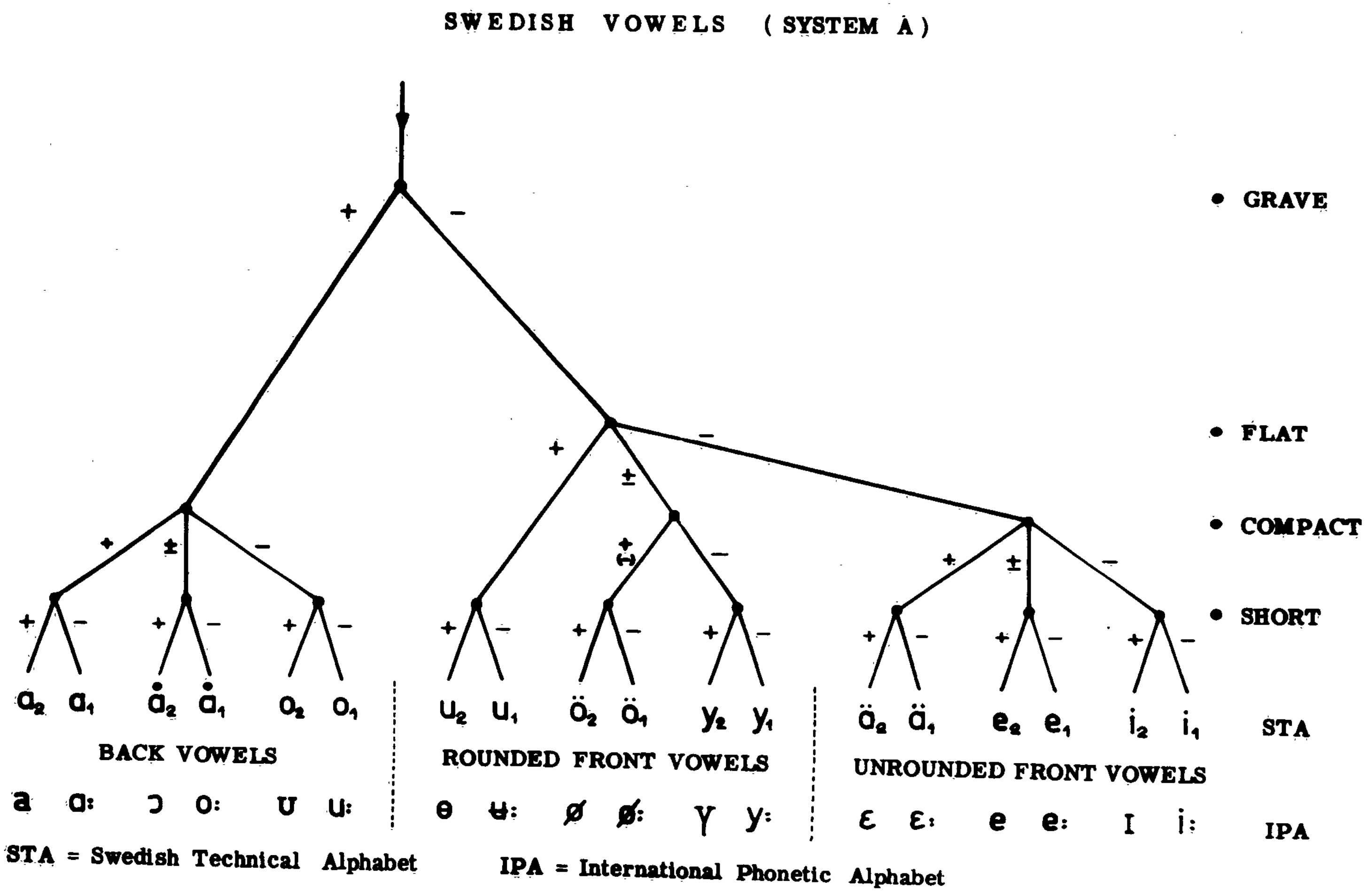
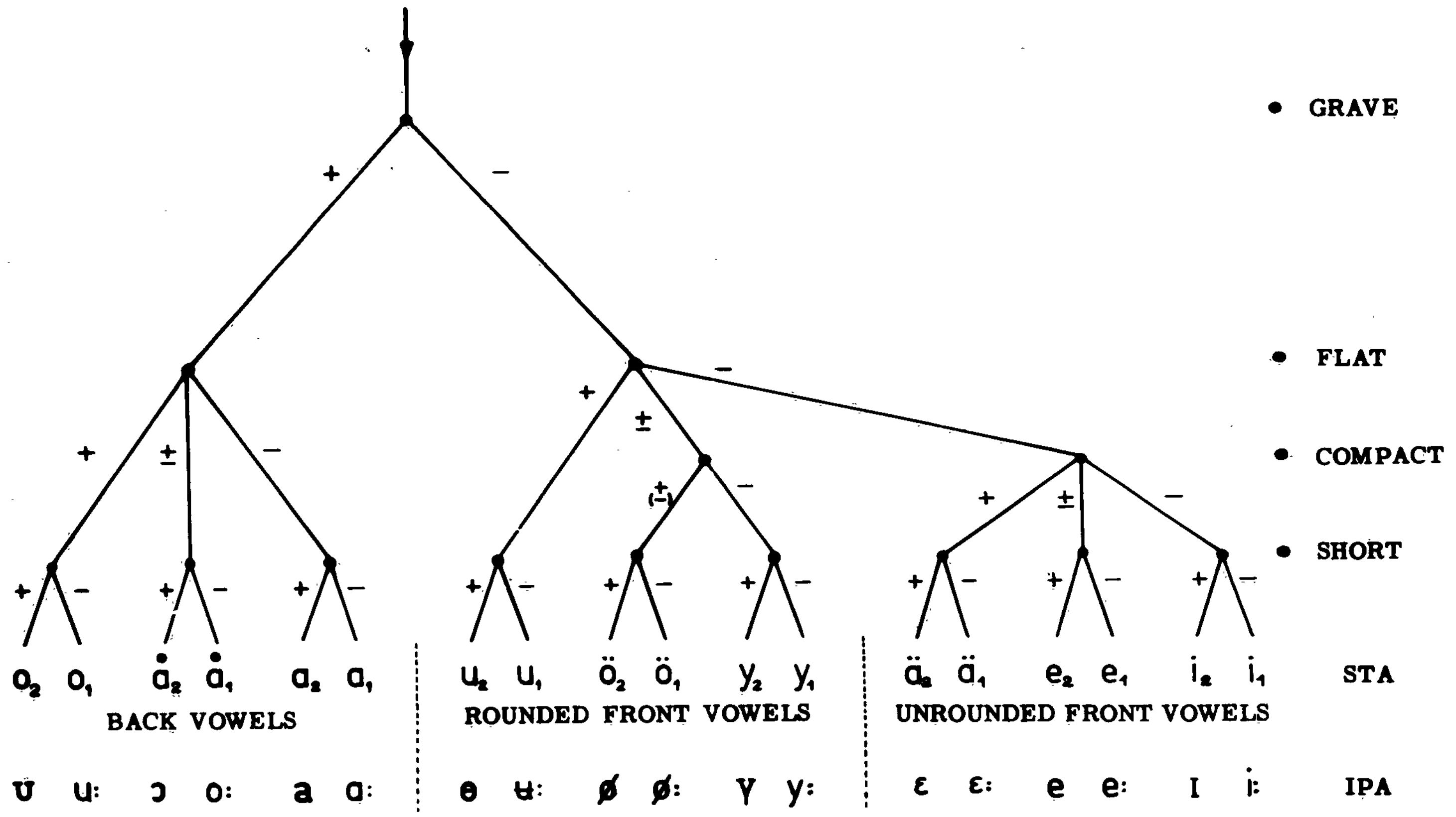


Fig. I-9 Distinctive feature coding of Swedish vowels. Back vowels separated in terms of compactness.

SWEDISH VOWELS (SYSTEM B)



STA = Swedish Technical Alphabet IPA = International Phonetic Alphabet

Fig. I-10 Distinctive feature coding of Swedish vowels. Back vowels separated in terms of flatness.

The relations within the rounded front vowels /y_1/ /u_1/ /$\ddot{o}_1$/ and /y_2/ /u_2/ /$\ddot{o}_2$/ are more complex. The phoneme /$\ddot{o}_1$/ is definitely compact compared with /y_1/ and /$\ddot{o}_1$/ is compact compared with /y_2/ but the phonemes /u_1/ and /u_2/ cannot be consistently specified by this feature. Thus /u_2/ is generally a more compact vowel than /$\ddot{o}_2$/ whereas /u_1/ is less compact than /$\ddot{o}_1$/. In general, however, the relation of /u_1/ to /$\ddot{o}_1$/ and /y_1/ or of /u_2/ to /$\ddot{o}_2$/ and /y_2/ is the same as that of /y_1/ and /$\ddot{o}_1$/ compared with /i_1/ /e_1/ and /$\ddot{a}_1$/. This is the motivation for the classification of /$\ddot{o}_1$/ and /y_1/ and similarly also /$\ddot{o}_2$/ and /y_2/ as ± flat.

Within the consonant system, Fig. I-11, the alveolar phonemos /rl/ /rn/ /rt/ /rd/ are opposed to the pure dentals /l/ /n/ /t/ and /d/ in terms of the flatness feature. From a distributional point of view, however, the alveolars may be regarded as the realization of a phoneme /r/ plus a following dental phoneme. This is also the case for an /rs/ as opposed to /s/ with the complication that [rs] stands in complimentary distribution to a quite similar sound labeled [sj] which is not the result of a fusion between an [r] and a dental. Thus /rs/ is the same phoneme as /sj/. The phoneme /sj/ is acoustically flat (lower frequency of main formant) both in relation to the compact (palatal) /tj/ and the non-compact acute /s/.

The following is a condensed summary of acoustic correlates of phonemic distinctions with special reference to the Swedish phoneme system*.

1. The vowel system

(1) Acute. An acute vowel has a higher F_2-F_1 then a corresponding non-acute (grave) vowel.

(2) Flat. A flat vowel has a lower sum (with possible weighting) $F_1+F_2+F_3$ than corresponding non-flat (plain) vowels.

(3) Compact. A compact vowel has a higher F_1 than a corresponding non-compact (diffuse) vowel.

(4) Short. A short vowel within the Swedish vowel system has a shorter duration and generally a spectrum with a higher F_1 and a more neutral formant pattern than a corresponding long vowel of the same context.

* The articulatory correlates are well established except for some specific details of the Swedish vowel system. For a general discussion of articulatory correlates, see earlier publications, e.g., G. Fant "Acoustic Theory of Speech Production".(5)

The long vowel is combined with a short consonant and vice versa. The short/long distinction is in all essentials identical with the lax/tense distinction described by Jakobson-Fant-Halle [1].

2. The consonant system

   (1) Vocalic. The F-pattern formants (F1 F2 F3) of frequencies F_1 F_2 F_3 respectively are more apparent and the overall intensity is higher in a vocalic than in a non-vocalic phoneme.

   (2) Consonantal. Consonantal phonemes are acoustically characterized by sound segments fulfilling one or both of the following conditions:
      a. The main energy is confined to other formants than F1 and F2.
      b. Low second formant intensity F2 and generally a low first formant frequency position F_1 compared with adjacent sound segment.

These aspects of the consonantal feature appear as a temporal contrast effect either in the type of spectrum or in terms of a rapid transition of the formant frequencies and the sound intensity.

There is an appreciable overlap in the definition of the consonantal feature and the non-vocalic feature which parallels the classification of most consonants as being both consonantal and non-vocalic. The phoneme /h/ is classified non-consonantal because of the lack of contrast of formant frequencies relative to an adjacent vowel but it is non-vocalic because of the lower intensity, especially in the first formant range. The liquids /l/ and /r/ have more vocalic formant patterns than other consonants and may thus be coded as +vocalic and +consonantal.

The syllable

The nucleus of a common syllable is a single vowel or a diphthong, the speech-wave correlate of which is a vocalic non-consonantal sound segment. An adjacent sound segment belonging to the same syllable is always consonantal and produces together with the vowel a temporal contrast owing to its lower intensity* or spectrum of a non-vocalic type. On account of the +vocalic feature liquids may be said

* It is probable that a proper frequency selective pre-emphasis in the intensity measuring procedure, favoring the frequency range below 3000 c/s, would invariably provide a larger intensity of a vowel than an adjacent consonant, e.g., fricative. Results from experiments on voicing detectors support this view.

to rank next after vowels in terms of "syllabicity". In a consonant cluster liquids always occur next to the vowel in conformity with the idea of a successive decay of syllabicity away from the syllable nucleus. A liquid /l/ or /r/ out of contact with a vowel would thus in itself constitute a syllabic nucleus. The /r/ possesses a greater syllabicity than /l/ on account of the greater compactness.

In unstressed syllables the vowel is of lower intensity, shorter duration, and possesses a formant pattern closer to that of a neutral vowel than in stressed syllables. The contrast between the vowel and adjacent consonants is reduced.

(3) Nasal. The formant structure of nasal sound segments has a reduced second formant intensity and possesses the typical qualities of the nasal murmur.

(4) Interrupted. Rapid onset or checking of the sound intensity combined with rapid formant transitions.

(5) Compact. The spectral energy of the sound segments of compact phonemes is concentrated to a more central location versus the main pitch of the immediately adjacent vocalic sound segments than non-compact (diffuse) phonemes*.

(6) Acute. Greater intensity of high-frequency formants than in non-acute (grave sounds).

(7) Flat. A shift down in the frequency location of formants retaining the general shape of the spectrum.

(8) Voiced. In Swedish consonants the voiced/voiceless opposition occurs in complimentary distribution with the lax/tense opposition. The common denominator is the relative lack or the shortness of duration of unvoiced segments of the speech wave. Examples are the burst and occlusion phases of voiced stops and the unvoiced segment of voiced (lax) continuants which are shorter in voiced than in unvoiced sounds.

* One suggested definition of main pitch is the second formant of a synthetic two-formant sound perceptually matching the quality of the sound segment. In first approximation this main pitch coincides with F_2 but is closer to F_3 or F_4 for high front vowels. In voiced stops and nasal consonants the front transitions carry a great part of the perceptually effective cues. The extent to which these cues may be included in the formulation above is not quite clear. In general the formant transitions and the consonantal sound segment constitute a compound stimulus.

Swedish consonants

		r	rl	l	ng	rn	n	m	g	k	rd	rt	d	t	b	p	sj =rs	j	tj	s	v	f	h
I	Vocalic	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
II	Consonantal	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−
III	Nasal				+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	
IV	Interrupted								+	+	+	+	+	+	+	+	−	−	−	−	−	−	
V	Compact	+	−	−	+	−	−	−	+	+	−	−	−	−	−	−		+	+	−	−	−	
VI	Acute					+	+	−			+	+	+	+	−	−				+	−	−	
VII	Flat		+	−		+	−				+	+	−	−			+	−	−	−			
VIII	Voiced								+	−	+	−	+	−	+	−		+	−		+	−	

Swedish vowels

		a_2	a_1	$å_2$	$å_1$	o_2	o_1	$ä_2$	$ä_1$	e_2	e_1	i_2	i_1	$ö_2$	$ö_1$	y_2	y_1	u_2	u_1
1	Acute	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+
2	Flat	(−)	(−)	(±)	(±)	(+)	(+)	−	−	−	−	−	−	±	±	±	±	+	+
3	Compact	+	+	±	±	−	−	+	+	±	±	−	−	+	+	−	−	−	−
4	Short	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−

G. Fant

(1) Jakobson, R., Fant, C.G.M., Halle, M.: "Preliminaries to speech analysis. The distinctive features and their correlates", M.I.T. Acoustics Lab. Techn. Rep. No. 13 (1952) 3rd printing.

(2) Fant, G.: "Modern instruments and methods for acoustic studies of speech", Proc. of the VIII International Congress of Linguists, Oslo University Press, Oslo 1958, p. 282-358; also publ. in Acta Polytechnica Scandinavica, Ph 1 (246/1958).

——— : "Acoustic analysis and synthesis of speech with applications to Swedish", Ericsson Technics No. 1, Vol. 15, 3-108 (1959).

(3) Fant, G.: "Phonetic and phonemic basis for the transcription of Swedish word material", Acta Oto-Laryngologica, Suppl. 116, 83-93 (1954).

(4) Fant, G.: "Descriptive analysis of the acoustic aspects of speech", paper presented at the 1960 Summer Symposia at Burg Wartenstein, Austria, "Comparative Aspects of Human Communication", September 4-10, 1960.

(5) Fant, G.: "Acoustic Theory of Speech Production", Mouton & Co., 's-Gravenhage 1960, 323 pp.

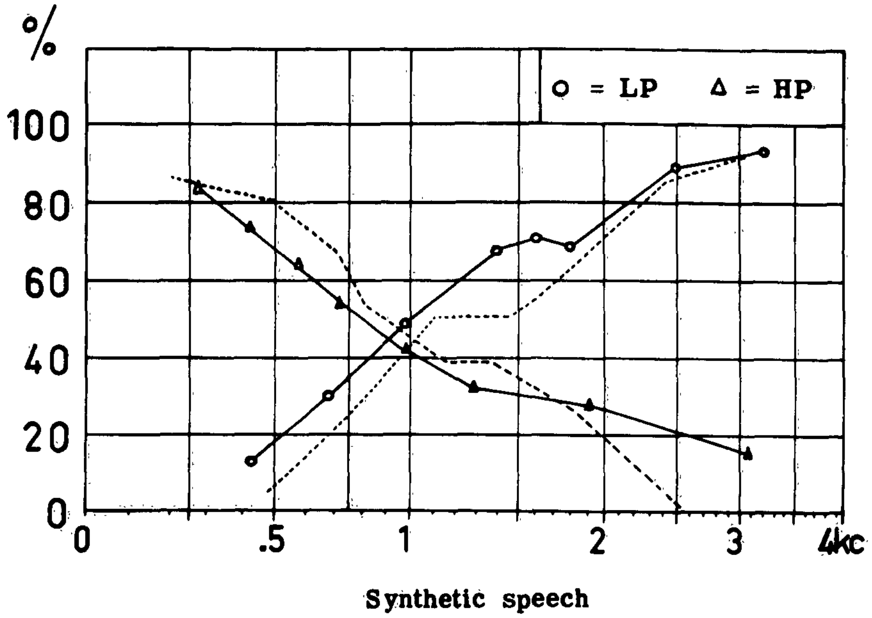
II. SPEECH SYNTHESIS AND SPEECH PERCEPTION

A. CONFUSION AMONG VOWELS FOLLOWING LOW-PASS AND HIGH-PASS FILTERING

Data from identification tests of HP- and LP-filtered Swedish vowels are being processed. Three male speakers and one synthetic speaker produced 14 different sustained vowels. A number of 9 subjects were used as listeners. The filters employed had very sharp cutoff, 1 dB per c/s. Mean results for a group of male speakers and for synthetic vowels are shown in Fig. II-1. It may be seen that the HP- and LP-curves cross at approximately 1000 c/s. Natural and synthetic speech provide similar results but the intelligibility of the synthetic speech was somewhat better than that of the mean of the human speakers.

The dashed curves pertain to a probabilistic theory of identification based on the number of formants within the pass band. Weighting factors of 2, 5, and 1 were applied to the first, second, and third formants respectively.

E.C. Carterette, A. Møller



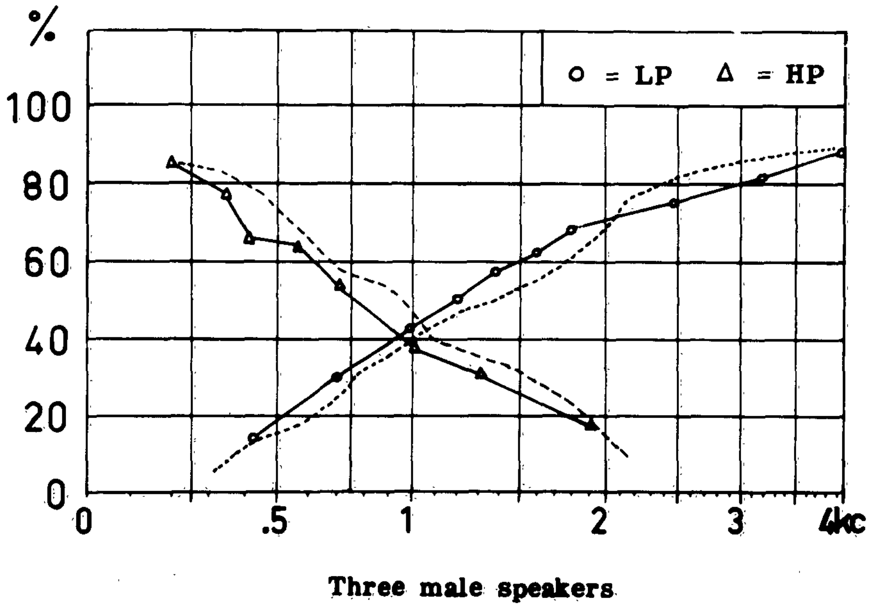


Fig. II-1 Per cent correctly identified Swedish vowels as a function of the cutoff frequency of low-pass and high-pass filtering. The top diagram pertains to synthetic speech and the lower diagram to the average of three human speakers. Dashed curves are derived from theoretical model.

B. SYNTHESIS INSTRUMENTATION

For pole-zero matching of speech samples by means of analog circuitry it is convenient to have units which represent a pole and a zero. The circuitry of Fig. II-2 has been developed for this purpose. The shunt arm containing a continuously variable inductance (feed-back amplifier variation) in series with a stepwise variable condenser determines the frequency of the zero. The pole is a function of the elements in both the series and the shunt arms. Bandwidths of the pole and the zero may be adjusted continuously. The unit is intended for the representation of any combination of a zero and pole of a frequency higher than the zero. The pole may be set high enough (with Q=1) in order to provide neglible contribution to the system function. Two additional units are under construction.

A few modifications of the synthesis instrumentation in OVE II for producing connected speech have recently been made. One is the zero circuit, see Fig. II-3, similar to that of Fig. II-2 but for the voltage control of the effective inductance and the use of integrators to provide the inductance. The advantage and the use of integrators to provide the inductance. The advantage gained by the inductance control is that overall reference gain of the unit does not change with the tuning. A new gate system, Fig. II-4, has also been designed in order to allow a more thump-free operation and a linear relation between control voltage and a dB signal level scale.

J. N. Holmes, M. Kringlebotn

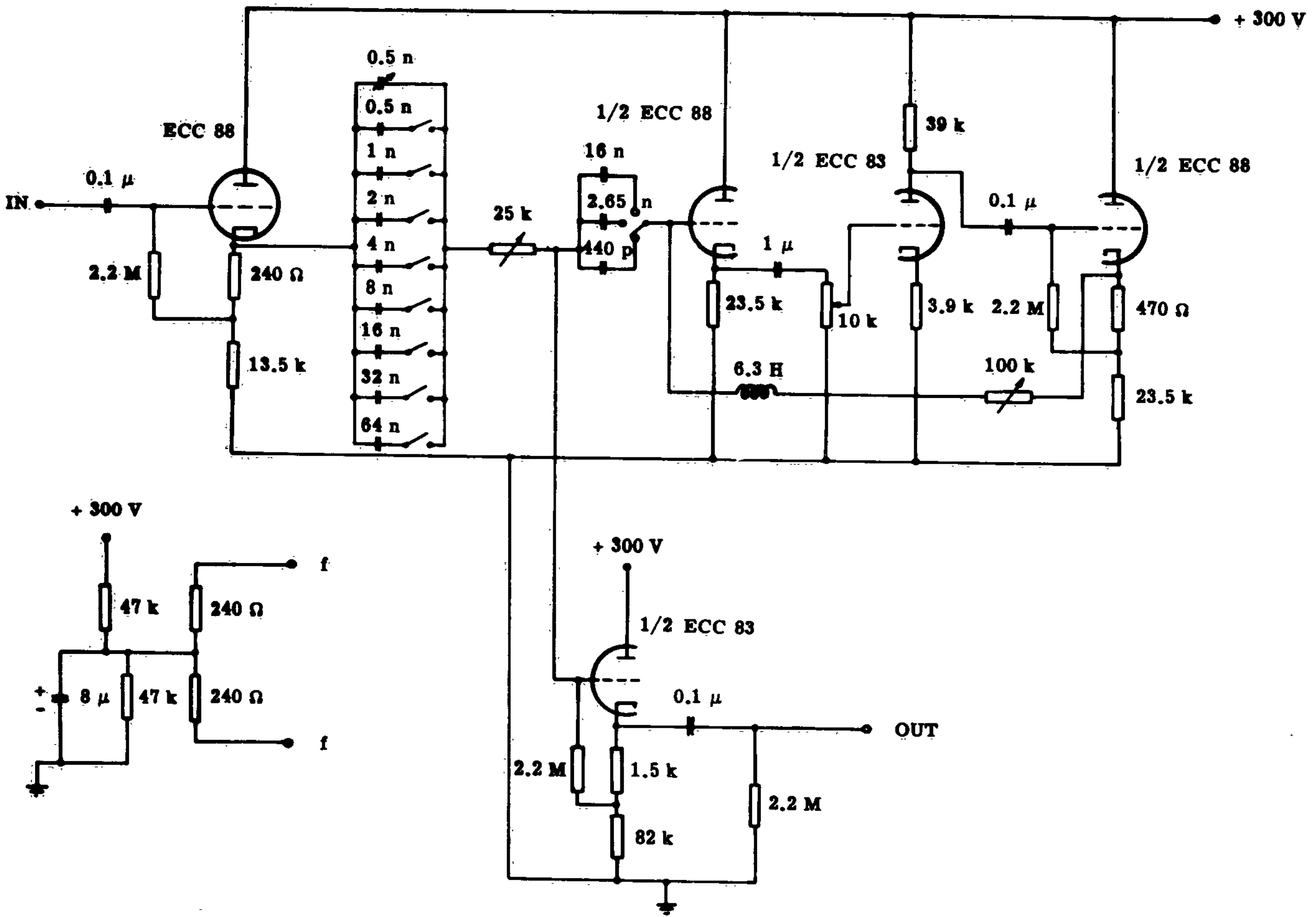
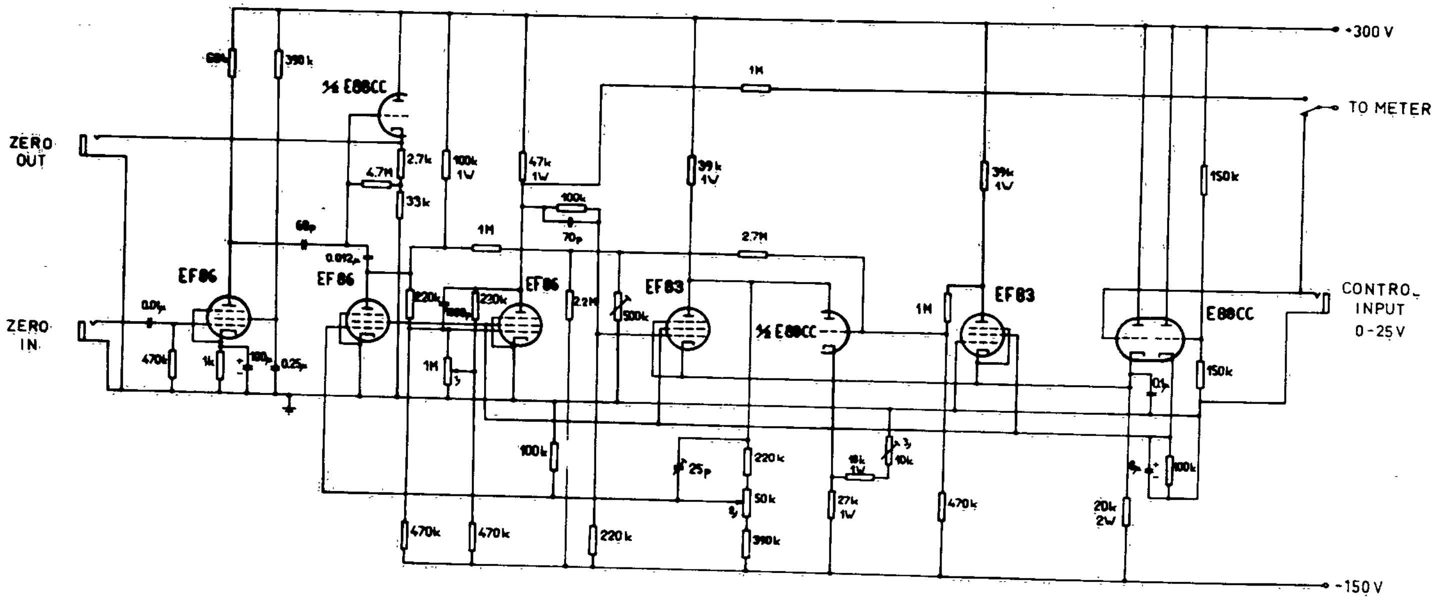


Fig. II-2 Circuit diagram of pole-zero unit for use in spectral matching by analog methods.



1 BANDWIDTH   2 DC LEVEL ADJUST   3 BALANCE ADJUST

Fig. II-3 Circuit diagram of a voltage controlled zero circuit for use in OVE II.

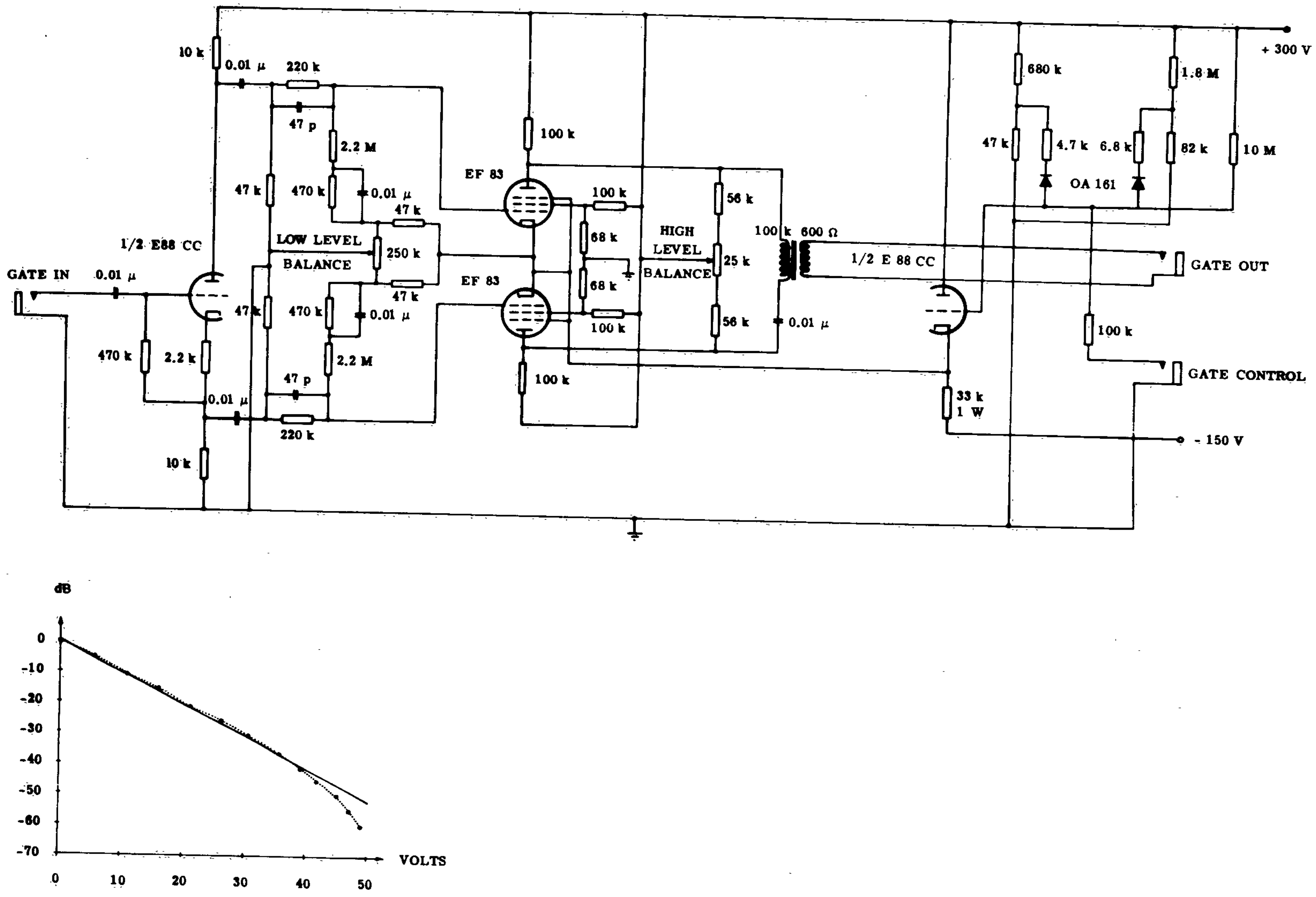


Fig. II-4 **Voltage controlled gate for OVE II providing linear variation on a dB-scale of signal level.**

III. SPEECH PRODUCTION

A. STUDIES OF NASALIZATION

Studies of synchronous records of cine-radiographic films and sound spectrograms have been performed on cleft palate subjects before and after operation and also on a control material of normal subjects. These studies have enabled a correlation between the degree of velo-pharyngeal opening and the nasalization element as seen in the spectrogram. The visual effects in a spectrogram of assimilated nasalization of vowel segments in immediate contact with a nasal consonant are small compared with those of the heavily nasalized speech of cleft palate subjects. A normal person can, however, to some extent simulate this pathological speech. Spectrographic pictures of simulated speech accordingly show a similarity to cleft palate speech.

Listening tests for assessment of speaker quality support the relative degrees of nasalization seen in the spectrographic and cine-radiographic material.

L. Björk, B. Nylén

Royal Inst. of Technology
Speech Transmission Lab.
Stockholm/Sweden

Summary
Rep.No. 1
AD______

Contract:
AF 61(052)-342
TN
TR

Monitoring Agency: AFCRC

FIELD: electronics,
speech information
processing

SPEECH ANALYSIS AND SYNTHESIS
C.G.M. Fant, January 31, 1961

ABSTRACT: The report summarizes progress in speech research during 1960. Among projects reported on are studies in speech sampling techniques and the design of spectrum sectioning devices. Results from studies of formant frequency measurements, formant tracking, pitch tracking, pole-zero matching, voice source regeneration, parametric synthesis of speech, and speech perception are discussed. Observations have been made on the production of nasalized sounds by means of cineradiographic and spectrographic methods.

USAF, European Office, ARDC, Brussels, Belgium

Royal Inst. of Technology
Speech Transmission Lab.
Stockholm/Sweden

Summary
Rep.No. 1
AD______

Contract:
AF 61(052)-342
TN
TR

Monitoring Agency: AFCRC

FIELD: electronics,
speech information
processing

SPEECH ANALYSIS AND SYNTHESIS
C.G.M. Fant, January 31, 1961

ABSTRACT: The report summarizes progress in speech research during 1960. Among projects reported on are studies in speech sampling techniques and the design of spectrum sectioning devices. Results from studies of formant frequency measurements, formant tracking, pitch tracking, pole-zero matching, voice source regeneration, parametric synthesis of speech, and speech perception are discussed. Observations have been made on the production of nasalized sounds by means of cineradiographic and spectrographic methods.

USAF, European Office, ARDC, Brussels, Belgium

Royal Inst. of Technology
Speech Transmission Lab.
Stockholm/Sweden

Summary
Rep.No. 1
AD______

Contract:
AF 61(052)-342
TN
TR

Monitoring Agency: AFCRC

FIELD: electronics,
speech information
processing

SPEECH ANALYSIS AND SYNTHESIS
C.G.M. Fant, January 31, 1961

ABSTRACT: The report summarizes progress in speech research during 1960. Among projects reported on are studies in speech sampling techniques and the design of spectrum sectioning devices. Results from studies of formant frequency measurements, formant tracking, pitch tracking, pole-zero matching, voice source regeneration, parametric synthesis of speech, and speech perception are discussed. Observations have been made on the production of nasalized sounds by means of cineradiographic and spectrographic methods.

USAF, European Office, ARDC, Brussels, Belgium

Royal Inst. of Technology
Speech Transmission Lab.
Stockholm/Sweden

Summary
Rep.No. 1
AD______

Contract:
AF 61(052)-342
TN
TR

Monitoring Agency: AFCRC

FIELD: electronics,
speech information
processing

SPEECH ANALYSIS AND SYNTHESIS
C.G.M. Fant, January 31, 1961

ABSTRACT: The report summarizes progress in speech research during 1960. Among projects reported on are studies in speech sampling techniques and the design of spectrum sectioning devices. Results from studies of formant frequency measurements, formant tracking, pitch tracking, pole-zero matching, voice source regeneration, parametric synthesis of speech, and speech perception are discussed. Observations have been made on the production of nasalized sounds by means of cineradiographic and spectrographic methods.

USAF, European Office, ARDC, Brussels, Belgium